Universidad de Córdoba
FACULTAD DE CIENCIAS
Departamento de Química Analítica

Mass spectrometry for the identification and quantitation of metabolomic biomarkers in clinical analysis

Espectrometría de masas para la identificación y cuantificación de biomarcadores metabolómicos en análisis clínico

Mónica Calderón Santiago

Tesis Doctoral
Córdoba 2014

TITULO: *Espectrometría de masas para la identificación y cuantificación de biomarcadores metabolómicos en análisis clínico*

AUTOR: *Mónica Calderón Santiago*

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

www.uco.es/publicaciones
publicaciones@uco.es

TÍTULO DE LA TESIS: ESPECTROMETRÍA DE MASAS PARA LA IDENTIFICACIÓN Y CUANTIFICACIÓN DE BIOMARCADORES METABOLÓMICOS EN ANÁLISIS CLÍNICO

DOCTORANDA: MÓNICA CALDERÓN SANTIAGO

INFORME RAZONADO DE LOS DIRECTORES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

La doctoranda Mónica Calderón Santiago ha sido una alumna brillante a lo largo de toda su formación. Durante el último año de licenciatura publicó un artículo de revisión en una de las revistas más punteras en Química Analítica (Trends in Analytical Chemistry, que ocupa el segundo puesto en el ranking de esta área) y consiguió una beca de Iniciación a la investigación del Programa Propio de la UCO. Terminada la licenciatura, cursó de forma brillante el máster en Química Fina Avanzada, en el que obtuvo excelentes calificaciones y obtuvo una beca-contrato de Formación de Profesorado Universitario, de la que viene disfrutando desde 2010 (convocatoria de 2009).

Durante el tiempo de disfrute de la beca-contrato, la doctoranda ha realizado investigación en el área de la metabolómica clínica, ha colaborado y supervisado investigación realizada por otros miembros del grupo del que forma parte, ha escrito capítulos de libros y ha presentado comunicaciones a congresos; todo lo cual se desglosa a continuación:

La investigación realizada ha abarcado:

1. El desarrollo de estrategias innovadoras con las que se contribuye a superar algunos puntos débiles de la metabolómica y que ha dado lugar a la publicación de 3 artículos publicados en revistas internacionales de alto índice de impacto.
2. La identificación de metabolitos biomarcadores de enfermedades (cardiovasculares y cáncer de pulmón) mediante el análisis global de biofluidos como la sangre o el sudor, lo que se ha reflejado en 4 artículos, de los cuales 1 ya ha sido publicado y 3 están en vías de publicación, y la solicitud de 1 patente internacional.
3. El análisis confirmatorio y la cuantificación de los biomarcadores con potencial clínico (ácidos tricarboxílicos, aminoácidos esenciales, fosfolípidos y péptidos), que ha resultado en 4 artículos de los cuales 3 han sido ya publicados y el restante ha sido aceptado para su publicación todos en revistas internacionales de alto índice de impacto.
4. La supervisión de investigación realizada por otros miembros del grupo ha dado lugar a 5 artículos, 4 de ellos publicados en revistas internacionales de alto índice de impacto.
5. La redacción de una revisión sobre el interés de las histatinas, que ha dado lugar a 1 publicación en TrAC Trends in Analytical Chemistry.
6. La contribución en 2 capítulos de libros multiautor *publicados por editoriales internacionales*.
7. La presentación oral o en cartel de 8 comunicaciones en congresos nacionales e internacionales.
8. Ha realizado una estancia de 3 meses en el Centro de Investigación Técnica VTT, de Finlandia, 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 Mónica Calderón Santiago.

Córdoba, 22 de mayo de 2014

Firma de los directores

Fdo.: M. D. Luque de Castro

Fdo.: F. Priego Capote

UNIVERSIDAD DE CÓRDOBA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA ANALÍTICA

ESPECTROMETRÍA DE MASAS PARA LA IDENTIFICACIÓN Y CUANTIFICACIÓN DE BIOMARCADORES METABOLÓMICOS EN ANÁLISIS CLÍNICO

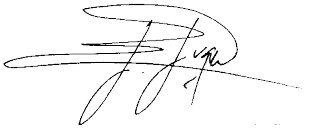
MASS SPECTROMETRY FOR THE IDENTIFICATION AND QUANTITATION OF METABOLOMIC BIOMARKERS IN CLINICAL ANALYSIS

Mónica Calderón Santiago

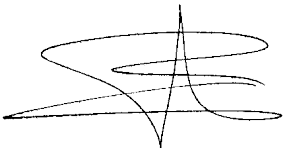
Córdoba, 2014

ESPECTROMETRÍA DE MASAS PARA LA IDENTIFICACIÓN Y CUANTIFICACIÓN DE BIOMARCADORES METABOLÓMICOS EN ANÁLISIS CLÍNICO

Los Directores,

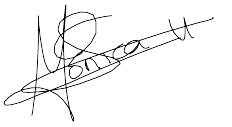


Fdo. María Dolores Luque de Castro
Catedrática del Departamento de
Química Analítica
Universidad de Córdoba



Fdo. Feliciano Priego Capote
Doctor Contratado Ramón y Cajal del
Departamento de Química Analítica
Universidad de Córdoba

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



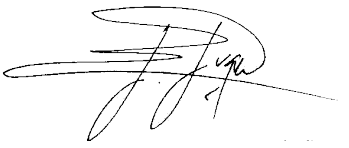
Fdo. Mónica Calderón Santiago
Licenciada en Química

María Dolores Luque de Castro, Catedrática, y **Feliciano Priego Capote**, Doctor Contratado Ramón y Cajal, ambos del Departamento de Química Analítica, Facultad de Ciencias, Universidad de Córdoba, en calidad de Directores de la Tesis Doctoral presentada por la Licenciada en Química Mónica Calderón Santiago, con el título "**Espectrometría de masas para la identificación y cuantificación de biomarcadores metabolómicos en análisis clínico**",

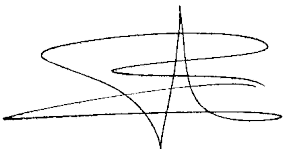
CERTIFICAN:

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 trabajos.

Y para que conste y surta los efectos pertinentes, expiden el presente certificado en Córdoba, a 2 de mayo de 2014



Fdo. María Dolores Luque de Castro



Fdo. Feliciano Priego Capote

Mediante la defensa de esta Memoria se pretende optar a la mención de **Doctora 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:
    - Prof. Dr. Jean-Michel Kauffmann, Lab. Instrumental Analysis and Bioelectrochemistry, Free University of Brussels, Bélgica.
    - Prof. Dr. Erwin Rosenberg, Vienna University of Technology, Institute of Chemical Technologies and Analytics, Austria.
2. Uno de los miembros del tribunal que ha de evaluar la Tesis pertenece a otro centro de Enseñanza Superior de otro país:
    - Prof. Dr. Reza Salek, Department of Biochemistry, University of Cambridge, Reino Unido.
3. La defensa de parte de esta Memoria se realizará en una lengua diferente a la materna: Inglés.
4. Estancia de tres meses en un centro de investigación de otro país:
    - Quantitative Biology and Bioinformatics (QBIX) department of the Technical Research Centre of Finland (VTT), directed by Tuulia Hyötyläinen and Marko Sysi-Aho, Espoo, Finlandia.

Agradecimientos

En primer lugar, quisiera dar las gracias a mis directores de tesis por su dedicación, su apoyo y su esfuerzo, que han hecho posible esta tesis. A la Catedrática María Dolores Luque de Castro, por confiar en mí desde el primer día, por transmitirme su ilusión por la investigación y por poner tantas ganas en que mi formación fuese lo mejor posible. Al Doctor Feliciano Priego Capote, por su infinita paciencia, por estar siempre dispuesto cuando lo he necesitado y por saber siempre encontrar una solución a los problemas. A ambos, mil gracias por todo el tiempo que habéis dedicado y por ser grandes profesionales y dignos ejemplos a seguir.

En segundo lugar, quisiera agradecer a todos mis compañeros no sólo la ayuda que me han aportado en el ámbito profesional, sino también esos buenos y no tan buenos momentos que hemos pasado juntos y que, finalmente, han hecho que estos años fuesen maravillosos. A Chema, mi primer "jefe", que aún hoy día sigue estando siempre ahí cuando lo necesito. A Mara, Sole, Bea y Miguel, porque aunque se hayan marchado y ya no podamos estar juntos todo lo que nos gustaría, nada podrá borrar los ratos de laboratorio que hemos pasado juntos. A Carlos, por ayudarme siempre que lo he necesitado y por esos inolvidables días en la "Chambre d'amour". A Auxi, mi compi de pilates y de otros muchos momentos de cánticos y risas. A Pili, por sus chistes y los ratos de gimnasio para "desestresarnos". A Angy, por las "carreras" por arroyo del moro y por dejarme que me meta con ella con cariño. A Vero, por ser como es, única, y por todos los momentos que hemos pasado juntas; espero que aún sean muchos los que nos queden por pasar. A Antonio, por dar ánimos cuando lo necesitas y por ser una gran persona. A María y Chuck, por los ratitos en coche de vuelta a Córdoba. A Carolina, María del Mar y Asun que aunque se han incorporado recientemente también me han hecho pasar buenos ratos. A Mayte, mi vecina, por tener excepcional sentido de humor y por hacer que de gusto estar en su compañía.
Gracias a todos, sabéis que "aunque no vine aquí para hacer amigos, siempre podréis contar conmigo".

También quiero dar las gracias a otros miembros del departamento con quienes he compartido ratos de feria, comidas de navidad y clases: Jose, M. José, Isa y María, Noelia, Merche, Laura, Guille, Carmen, Godoy,... No puede faltar mi agradecimiento a Diego y Juani, que han estado siempre disponibles cuando he necesitado algo.

I also would like to tkanks Tuulia Hyötyläinen for welcoming me into her research group, as well as the members of her group: Päivi Pöhö, Maarit Kivilompolo, Sirkku Jäntti, Isabel Bondia, Benedicte,... for making me easier the three months that I stayed in Finland.

A mis padres, porque soy lo que soy gracias a ellos, por haberse sacrificado y haberme animado para que pudiese llegar donde quisiese. A mis hermanos y mi sobrina, por ayudarme a evadirme y pasar buenos ratos al llegar a casa. A Fernando, por haberme apoyado y animado en mis momentos de flaqueza y por saber entenderme como nadie, gracias por estar a mi lado cada día.

“Soy de las que piensan que la ciencia tiene una gran belleza. Un científico en su laboratorio no es sólo un técnico: es también un niño colocado ante fenómenos naturales que le impresionan como un cuento de hadas.”

Marie Curie

ÍNDICE

Index

OBJETIVOS

El *objetivo global* de esta investigación fue desarrollar estrategias de análisis metabolómico global y orientado para la identificación y cuantificación de metabolitos con potencial como biomarcadores en análisis clínico, para lo que se utilizó una técnica de detección clave como es la espectrometría de masas — que destaca por su gran versatilidad y excelentes niveles de sensibilidad, selectividad, precisión, exactitud y resolución. Se trata pues de un objetivo que engloba una disciplina en auge como es la metabolómica y un área de investigación como es la identificación de biomarcadores en análisis clínico. Este objetivo se dividió a su vez en tres *objetivos generales* en función de la temática del estudio en cuestión: Objetivo 1, desarrollar estrategias innovadoras en análisis metabolómico global con aplicación en clínica: (a) en una muestra escasamente estudiada como es el sudor, (b) en la preparación de la muestra y, (c) en la detección por espectrometría de masas para conseguir una mejora significativa en los métodos globales. Objetivo 2, identificar metabolitos con potencial como biomarcadores mediante el análisis global de biofluidos para su aplicación en estudios clínicos nutricionales o en el diagnóstico de enfermedades como la aterosclerosis y el cáncer de pulmón. Objetivo 3, optimizar métodos de análisis orientado confirmatorio y cuantitativo (absoluto y relativo) para la determinación de compuestos con potencial como marcadores en biofluidos y su aplicación en el diagnóstico de enfermedades.

De estos tres objetivos generales derivaron los siguientes *objetivos concretos*:

i) Realizar un estudio de optimización para la evaluación en análisis metabolómico de un biofluido obtenido de forma no invasiva como es el sudor. Se pudieron así comparar diferentes alternativas de preparación de muestra con las que proponer un método adecuado para la utilización analítica de este biofluido. Este estudio permitió la identificación mediante análisis por espectrometría de masas de

metabolitos presentes en sudor del que existe escasa información sobre su composición. Esta investigación ha dado como resultado un artículo (Capítulo 1) que recoge los aspectos derivados del análisis de esta muestra poco frecuente.

ii) Evaluar la utilidad de una plataforma analítica totalmente automatizada basada en el acoplamiento en línea de la extracción en fase sólida (SPE) y la cromatografía de líquidos con la espectrometría de masas en tándem (LC–MS/MS). Para ello se usó una gran diversidad de materiales sorbentes y la espectrometría de masas de alta resolución mediante un detector híbrido cuadrupolo–tiempo de vuelo (QTOF), para su aplicación al análisis global de suero humano, según recoge el Capítulo 2.

iii) Demostrar una nueva faceta de la detección por espectrometría de masas mediante una modalidad conocida como Fraccionamiento en Fase Gaseosa (GPF) para su implementación en análisis global, tal como muestra el Capítulo 3.

Un objetivo general y amplio como es la identificación de metabolitos biomarcadores mediante el análisis global de biofluidos se ha dividido en función del área de aplicación: Intervenciones dietéticas y diagnóstico de enfermedades como la aterosclerosis y el cáncer de pulmón, lo que ha dado lugar a una división en los siguientes *objetivos concretos*:

iv) Estudiar mediante análisis metabolómico global basado en LC–QTOF las diferencias existentes en el perfil metabólico del suero de individuos sometidos a cuatro intervenciones dietéticas basadas en diferente contenido energético y graso (Capítulo 4). De esta forma, se pretende establecer mediante análisis metabolómico diferencias metabólicas asociadas a diferentes patrones dietéticos.

v) Elucidar las diferencias metabólicas entre individuos con aterosclerosis diagnosticados con angina estable, infarto de miocardio sin elevación del segmento ST e infarto agudo de

miocardio (Capítulo 5). El papel de la espectrometría de masas de alta resolución es clave para la identificación de metabolitos con contribución significativa a la discriminación.

vi) Buscar un panel formado por metabolitos con potencial como biomarcadores para monitorizar individuos diagnosticados con aterosclerosis y predecir un episodio de infarto agudo de miocardio. La inclusión de factores clínicos en la búsqueda de los modelos predictivos mediante la combinación de marcadores estadísticos es clave en numerosos estudios, tal y como se detalla en el Capítulo 6.

vii) Aplicar el protocolo derivado de i) a la discriminación de individuos diagnosticados con cáncer de pulmón frente a individuos sanos (incluyendo fumadores y no fumadores). La utilización de un biofluido no invasivo como el sudor y el desarrollo de un panel de metabolitos con potencial marcador son la base para la consecución de este objetivo planteado y resuelto en el Capítulo 7.

El objetivo general 3, relacionado con el análisis orientado para el desarrollo de métodos de análisis confirmatorio y cuantitativo de compuestos con potencial como marcadores clínicos, tuvo como campo de aplicación el análisis de compuestos polares tales como ácidos tricarboxílicos y aminoácidos esenciales, compuestos lipídicos como fosfolípidos y un péptido de interés clínico como la catelicidina. Con estas premisas, los *objetivos concretos* fueron:

viii) Utilizar el perfil de fosfolípidos en suero humano para la búsqueda de un panel con potencial predictivo para discriminar individuos diagnosticados con aterosclerosis. La selección de una etapa de SPE para la separación selectiva de los fosfolípidos y la espectrometría de masas de alta resolución constituyen dos herramientas claves de esta investigación junto al tratamiento de datos diseñado para configurar el panel de fosfolípidos marcadores, según se muestra en el Capítulo 8.

ix) Establecer diferencias en concentración de aminoácidos esenciales en suero humano mediante SPE–LC–MS/MS en pacientes con aterosclerosis afectados por angina estable o por infarto agudo de miocardio. Para ello, la utilización de la configuración instrumental basada en un detector de triple cuadrupolo aporta una herramienta óptima en análisis cuantitativo orientado con posibilidad de analizar suero humano sin tratamiento previo (Capítulo 9).

x) Determinar las diferencias metabólicas en pacientes con determinados factores de riesgo cardiovascular respecto a la concentración en suero de metabolitos implicados en el ciclo de los ácidos tricarboxílicos. Identificar interacciones de factores de riesgo con la presencia de lesiones coronarias fue un objetivo adicional planteado, que también se recoge en el Capítulo 10.

xi) Desarrollar un método basado en el análisis directo de suero humano para el análisis confirmatorio y cuantitativo de péptidos de interés clínico como es la catelicidina. La aplicación de este método al análisis de suero de pacientes de la unidad de cuidados intensivos puede ser de interés, tal como se describe en el Capítulo 11.

La formación de la futura doctora, que constituye el *objetivo último* de toda tesis doctoral, ha incluido el máster en "Química Fina" con el número de créditos correspondientes, así como todas las etapas necesarias para cumplir los requi-sitos exigidos para optar a la mención de Doctorado Internacional. En paralelo con lo anterior y con la investigación recogida en la parte principal de la Memoria, se ha pretendido una formación más amplia de la doctoranda mediante la realización de otras actividades que se recogen como anexos, a saber:

Anexo I: Investigación simultánea con la de la Tesis y supervisión de trabajos realizados por otros miembros del Grupo, que han dado lugar a 5 artículos publicados o en fase de publicación.

Anexo II: Revisión bibliográfica sobre el interés y el potencial de la determinación de histatinas (péptidos antimicrobianos) en muestras biológicas.

Anexo III: Capítulos de libro sobre temas relacionados o no con la Tesis, como las aplicaciones de la metabolómica o las técnicas empleadas para determinar la acción de una bacteria sobre ciertos metabolitos en leche.

Anexo IV: Patente de un panel de marcadores para el diagnóstico de cáncer de pulmón mediante análisis del sudor humano.

Anexo V: Comunicaciones orales y en cartel en 8 conferencias nacionales o internacionales.

OBJECTIVES

The ***basic objective*** of the research in this book was to develop strategies based on targeted and untargeted metabolomics analysis for identification and qualitative/quantitative analysis of metabolites with biomarker capability in clinical analysis by using mass spectrometry as analytical technique —because of its versatility and excellent levels of sensitivity, selectivity, precision, accuracy and resolution. This is, therefore, an objective that encompasses the current impact of metabolomics and research on the search for clinical biomarkers. This objective was divided into three ***general objectives*** according to the different topics in this research: Objective 1, to develop innovative strategies in untargeted metabolomics analysis with clinical application: (a) by characterization of one sample scarcely studied as does human sweat, (b) in sample preparation and, (c) in detection by mass spectrometry in order to enhance the detection capability of untargeted methods. Objective 2, to identify metabolites with biomarker potential by global analysis of both a conventional biofluid such as serum and a unconventional biofluid such as sweat for application in dietetic clinical studies or in the diagnostic of diseases such as atherosclerosis or lung cancer. Objective 3, to optimize confirmatory and quantitative (absolute and relative) methods for targeted analysis of metabolites with biomarker potential in biofluids and their application for diagnostic of pathological states.

From the general objectives derived the following ***concrete objectives***:

i) To perform an optimization study for assessment in metabolomics analysis of a non-invasive biofluid such as sweat. Thus, it was possible to compare different sample preparation alternatives to propose a suited method for analytical application of this biofluid. This study allowed identification of metabolites present in this biofluid by mass spectrometry, which provided an added value to this study supported on the lack of publications dealing with the characterization of

human sweat. The result was a research article containing results obtained by analysis of this sample (Chapter 1).

ii) To evaluate the usefulness of an automated analytical platform based on on-line coupling of solid phase extraction (SPE) and liquid chromatography with mass spectrometry (LC–MS/MS) in untargeted analysis. For this purpose, a great diversity of sorbent materials and high resolution mass spectrometry by a hybrid quadrupole–time of flight (QTOF) detector were two crucial aspects in this research. The application area was the global analysis of human serum, as Chapter 2 details.

iii) To demonstrate the potential of Gas-Phase Fractionation (GPF) as mass spectrometry detection mode in global metabolomics analysis (see Chapter 3).

A wide general objective as the identification of biomarker metabolites through global analysis of biofluids has been divided into the following *concrete objectives* according to the application field: dietetic interventions and diagnostic of diseases such as atherosclerosis and lung cancer:

iv) To study, by untargeted metabolomics analysis based on LC–QTOF MS/MS, the differences in the metabolic profile of serum from individuals subjected to four dietetic interventions characterized by the energetic content and fatty composition (Chapter 4). The study was oriented to find metabolic differences associated to the different dietetic patterns through metabolomics analysis.

v) To elucidate metabolic differences among atherosclerotic individuals diagnosed with: stable angina, non-ST elevation myocardial infarction (NSTEMI) and acute myocardial infarct (Chapter 5). The role of high resolution mass spectrometry is crucial for identification of metabolites with significant contribution to explain the observed discrimination.

vi) To find a panel formed by metabolites with biomarker potential to monitor individuals diagnosed with atherosclerosis and predict a possible episode of myocardial infarct. The inclusion of clinical factors in the search for predictive models by iterative combination of statistical biomarkers is key in this type of studies, as Chapter 6 describes.

vii) To apply the protocol optimized in i) to discriminate individuals diagnosed with lung cancer against healthy individuals (including smokers and non-smokers). The utilization of a non-invasive biofluid and development of a metabolites panel with biomarker potential are the basis to achieve this objective developed in Chapter 7.

General objective 3, related to targeted analysis and development of confirmatory and quantitative analytical methods for determination of metabolites with clinical marker capability, was focused on the analysis of polar compounds such as carboxylic acids and essential amino acids, lipids such as phospholipids, and a peptide with clinical interest such as cathelicidin. With these premises, the *concrete objectives* were:

viii) To use the phospholipids profile of human serum to find a panel of compounds with predictive capability to discriminate indi-viduals diagnosed with atherosclerosis. The selection of a SPE step for isolation of phospholipids and the selection of high-resolution mass spectrometry constitute two basic tools in this research, together with the data treatment designed to propose the panel of marker phospholipids, as can be seen in Chapter 8.

ix) To differentiate the concentrations of essential amino acids in human serum by SPE–LC–MS/MS in patients with atherosclerosis also affected either by stable angina or acute myocardial infarction. The use of analytical equipment involving a triple quadrupole detector consti-tutes an optimal tool for quantitative targeted analysis,

capable of analyzing human serum without sample preparation (Chapter 9).

x) To establish metabolic differences among patients with given cardiovascular risk factors related to serum concentration of metabolites involved in the tricarboxylic acids cycle. To identify interaction of risk factors with coronary lesions was an additional objective also included in Chapter 10.

xi) To develop a method based on direct analysis of human serum for confirmatory and quantitative analysis of peptides of clinical interest as is the case with cathelicidin, as described in Chapter 11.

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

- Annex I: Research simultaneous with that of the Thesis and supervision of research carried out by other members of the Group, which have provide 5 articles published or in some step prior to publication.
- Annex II: Review about the interest of the determination of a certain group of antimicrobial peptides, called histatins, in biological samples.
- Annex III: Book chapters on subjects related and non-related to the Thesis, such as the applications of metabolomics or the techniques employed to study the effect of a bacteria on certain metabolites in milk.

- Annex IV: A patent of a markers panel for the diagnosis of lung cancer by sweat analysis.
- Annex V: Oral and poster communications in 8 national or international meetings.

Introducción

Introduction

This introduction section is intending to offer an overview of the two main topics considered in this research that constitutes this PhD Book: Metabolomics and qualitative and quantitative analysis of biomarkers.

1. Location, definition and characteristics of metabolomics

From the Latin suffix "ome", the term "omics" means mass or many and is used to differentiate studies which involve large number of measurements per endpoint. The measured parameters can be DNA, mRNA, proteins, and metabolites, giving place to primary omics (genomics, transcriptomics, proteomics and metabolomics, respectively), or they can be specific families such as lipids, glycans, toxics, foreign substances, or even interactions among them, providing the secondary omics (lipidomics, glycomics, toxicomics, xenometabolomics, and interactomics, respectively), all them as parts of primary omics. The evolution and discovery of omics complementarity yielded the term systems biology, coined by Nicholson and Wilson [1], and also expressed as high-dimensional biology, global systems biology or integrated systems biology. The term intended, in principle, to integrate multivariate biological information to better understand gene–environment interactions. Figure 1 schematizes both the traditional central dogma of molecular biology, where the flow of information goes from genes to transcripts to proteins, finishing with the site where enzymes act on metabolism (Figure 1.A), and also the omics organization, where the flow of biological information ends in the metabolites which finally is reflected in the phenotype (Figure 1.B).

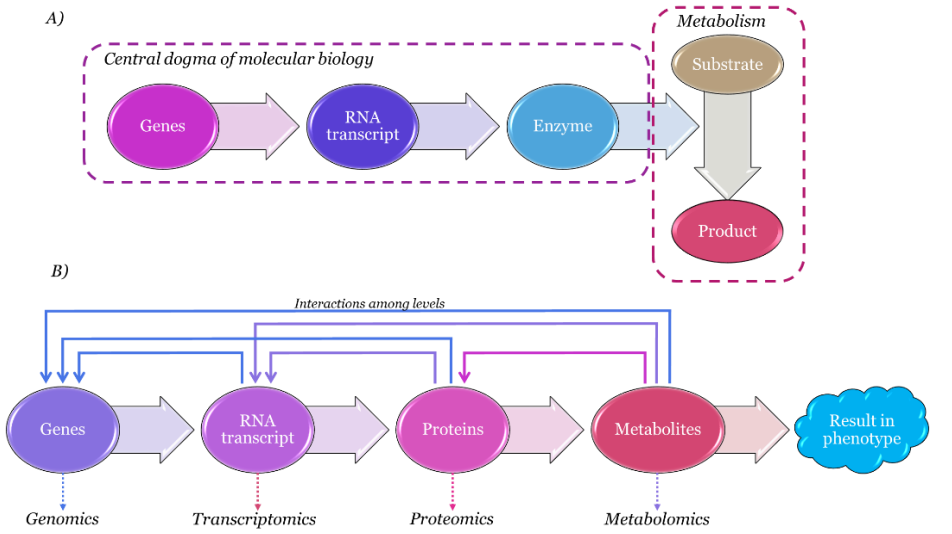


Figure 1. A) Traditional central dogma of molecular biology. **B)** General scheme of systems biology.

Through the study of metabolomics it is possible to define the current status of the system, as a result of environmental conditions and genetic potential, that is more difficult to be predicted by genomic and/or proteomic characterizations. For this reason, regardless of the sequential relationship in which the omics are introduced, metabolomics provides information directly in line with the new approach of systems biology that considers living systems as dynamic and complex and assumes that their behavior originates from interactions; therefore, it is difficult to make predictions when exclusively considering the properties of individual parts [2–4].

In any case, the subject of metabolomics is the metabolome, which includes all organic substances naturally occurring from the metabolism of the studied living organism, also including xenobiotics (i.e., chemicals that are present in an organism, but not produced by it) and their biotransformation products. Polymerized structures such as proteins and nucleic acids are excluded from the metabolome. Actually, one may consider that a metabolite correspond to any organic compound that does not directly come from gene expression. According to this view, small peptides —synthesized through an enzymatic

process and forming part of a metabolic pathway, as is the case of tripeptide glutathione— are considered metabolites.

Endogenous metabolites can be classified into primary and secondary metabolites. The former are ubiquitous (*i.e.,* they may occur in bacterial, plant, and animal kingdoms) and are directly involved in processes essential to life, such as growth, development, and reproduction. This is, for example, the case for amino acids or intermediates of glycolysis. At the opposite, secondary metabolites have a restricted distribution (often at the species level) and are synthesized for a particular biological function. This is, for example, the case with alkaloids in plants or hormones in mammals. At present, approximately 2.900 endogenous metabolites can be identified, but if we consider their total number, estimated at 6.500 molecules, this is lower than the 25.000 genes, the 300.000 tRNAs and the more than 10^6 proteins present in the human organism [5,6].

Despite the term metabolomics has been traditionally attributed to Fiehn, who in 2002 defined this discipline as "an overall and exhaustive analysis by which all the metabolites in a biological system are identified and quantified" [7], a previous definition by Oliver (in 1998) established metabolomics as "the analysis of the metabolome under a given biological condition and referred to a high-scale detection and quantitation of metabolites in biological media" [8]. Presently, the definition of metabolomics as a more or less complex analysis is very poor and a wider and well documented definition as "the discipline that provides comprehensive and sistematic information on temporal changes in the profiles of metabolite levels in biofluids and tissues, which can arise from control by the host genome, extended genomes and effect of other enviromental or promoted factors" is better adjusted to metabolomics. Implicitely included in the definition, without expliciting their concourse are transcripts and proteins, thus involving the top-down pathways established by the central dogma of molecular biology, and the bottom-up positive and negative feedback control established by systems biology (Figure 1).

Distinction between metabolomics and metabonomics, the two terms used to design the discipline, has been differently interpreted by the authors working in this field. While some of them consider metabolomics as the complete analysis of plants metabolome, and metabonomics as the analysis of the metabolic reaction to medication, environment and/or diseases; others establish that the goal of metabolomics is holistic quantitative and qualitative analysis of all metabolites in a given biological system, whereas the goal of metabonomics is the comparison of metabolite levels in a given system in response to a particular stimulus of treatment [9,10]. Nevertheless, one or the other distinction is each time more infrequent and the term metabolomics tends to be used in all instances.

The present state of metabolomics endows this discipline with positive and negative technical characteristics that are in the way of being in depth studied by researchers in the field for even improving the former and, of course, overcoming the latter. Therefore, positive characteristics of metabolomics that promote its use and establish a difference with other omics are as follows:

- The existence of robust and stable analytical platforms for a number of metabolites, as a result of the research on this area during the last two decades of the twenty century, before omics as such appeared.
- The excellent analytical and biological precision of the existing platforms that allows their accurate use.
- The analysis cost per sample and analyte, which is, in general, low as compared with genomics and transcriptomics analyses. This fact facilitates application to a high number of samples, and extensive studies in cohorts with a significant biological variability; thus providing wider and well-supported information.
- The analysis time, much shorter than that required in genomics, transcriptomics and proteomics, especially for fingerprinting, footprinting or qualitative metabolomics profiles, which are obtained in few min.

- The capability to integrate data from different techniques (integrated metabolomics) applied to samples from different body compartments.

Negative or non-resolved aspects of metabolomics are as follows:

- The necessity for multiple analytical platforms as a result of the enormous variety of metabolites that ranges from non-polar lipids to ionic compounds; from relatively high molecules such as cholesterol to very small ions such as ammonium.
- The wide range of concentrations within which metabolites appear in the organisms, which encompasses more than nine orders of magnitude (from pmol to mmol), and makes mandatory the use of dilution and preconcentration steps, as well as analytical equipment with different dynamic range.
- The absence of enough standardized methods, as a consequence of being the last of the primary omics.
- The number of metabolites constituting the metabolome, which in a vegetal system can reach 200.000 (in a human individual lower than 10.000); and the diversity within a single family.

2. Complementarity of omics

Despite metabolomic approaches have been trialing somewhat behind transcriptomics and proteomics approaches, key metabolomics contributions demonstrated more than a decade ago, and continue demonstrating, the complementary and sometimes decisive role of metabolomics in the omics-world. For example:

- by demonstrating that metabolic pathways can be determined kinetically by monitoring correlative behaviors between only a handful of metabolites [4],

- by proofs of functionality in the identification of signal transduction cascades [5],
- by using the analysis of intracellular metabolites to reveal phenotypes for mutations of proteins active in metabolic regulation [6],
- by quantifying the change of several metabolite concentrations relative to the concentration change of one selected metabolite to reveal the site of action, in the metabolic network, of silent genes [7],
- by developing metabolite profiling studies to identify hyperbolically related pairs of metabolites that were consistent with known feedforward and feedback mechanisms of regulation [8].

Metabolomics, as the complement to transcriptomic and proteomics, has its own advantages as compared with them, such as that there are fewer metabolites than genes and proteins (as example, the human genome contains 25.000 genes, the expression and alternative splicing of the mRNAs indicate that human may be able to produce 10^6 different proteins, while a recent prediction in 2013 suggests that there are only 40.000 metabolites (6.500 in humans) [9]); thus reducing processing complexities. Additionally, the technology involved in metabolomics is generic, as a given metabolite —unlike a transcript or protein— is the same in every organism that contain it.

One other benefit is that, although some environmental perturbations or genetic manipulations may not cause changes in transcriptomic and proteomic levels, they will have significant effects on the concentrations of numerous individual metabolites due to the "downstream" character of metabolomics. Even more, metabolomics lends itself readily to functional genomic and other analyses, as changes in the cell physiology as a result of gene deletion or over-expression are amplified, both in theory [10] and in practice [11], through the hierarchy of the transcriptome and the proteome and are, therefore, more easily measurable through the metabolome, even when changes in metabolic fluxes are negligible. Finally, the dynamic levels of metabolites in organisms must reflect the exact

metabolic phenotypes under different cultural and genetic conditions due to the direct interaction between metabolome and phenotype.

Nevertheless, there is in metabolomics a typical complexity derived from the chemical diversity of the metabolome due to the large number of different families of metabolites that hinders to obtain a complete view of the metabolome by using a unique analytical method. Therefore, the combination of orthogonal methods is practically mandatory. Additionally, there is a number of intrinsic and extrinsic factors affecting metabolic pathways, as Figure 2 shows. In addition, measurement of a complete set of metabolites not only requires rapid quenching of metabolism, in the case of intracellular metabolites, but also the development of a method suitable for detecting and quantifying large numbers of metabolites depending of the intrinsic and extrinsic factors of the given individual. It is worth emphasizing the rapid timescale for the turnover (the turnover times is approximately equal to the concentration of the metabolite divided by the flux through the pathway on which it is a member) of an intracellular metabolite, which in microbial systems can be under 1 s, even for metabolites at mM concentrations.

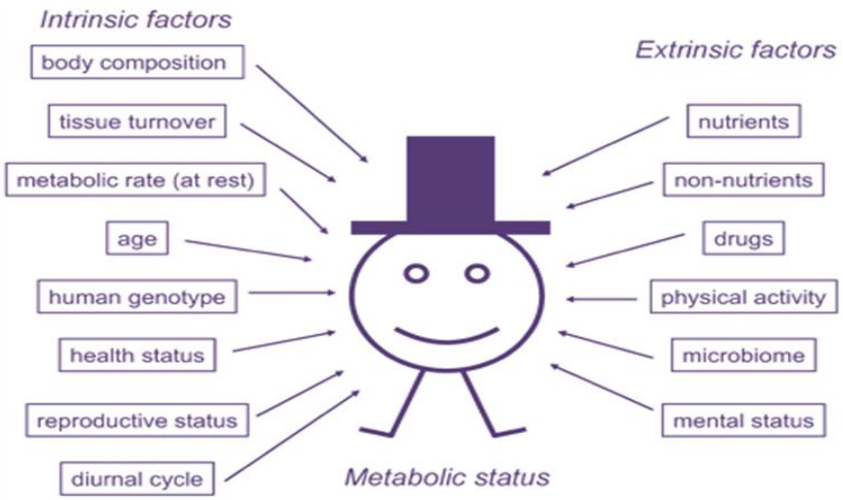


Figure 2. Intrinsic and extrinsic factors that affect the human metabolic status and can be studied using metabolomics.

3. Analytical strategies in metabolomics and their main objectives

Metabolomics analysis encompasses different strategies which depend on the information required from the system under study [12], namely: (a) targeted analysis [13], which aims at qualitative and quantitative study of one or, more frequently, a small group of chemically similar metabolites; (b) global metabolomics profiling [14], which allows detection of a broad range of metabolites by using a single analytical platform or a combination of complementary analytical platforms —mainly based on nuclear magnetic resonance (NMR) or mass spectrometry (MS), coupled to gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE)— to obtain a comprehensive profile of the metabolome; (c) metabolomics fingerprinting [15], a high throughput, fast methodology for analysis of biological samples that provides fingerprints for sample classification and screening; and, (d) metabolomics footprinting [16], which aims at the study of metabolites in extracellular fluids, also known as exometabolome or secretome. Obviously, the complexity of sample preparation for each strategy is different increasing from (c) to (a).

The most recent trend leans to simplification by establishing distinction only between targeted and untargeted analysis, including profiling, which require quite different sampling and sample preparation approaches [17]. Each of these strategies has its own inherent advantages and disadvantages, but they can be highly complementary when used in combination.

Untargeted metabolomics is the attempt to measure all the analytes in a sample, including chemical unknowns. Due to its comprehensive nature, analysis of untargeted metabolomic data must be conducted with advanced chemometric techniques, such as multivariate analysis, designed to reduce the extensive data sets generated into smaller and more manageable signals. These unidentified signals then require annotation using either available experimental libraries, structural elucidation by *in silico* fragmentation tools or experimental identi-

fication with standards. From the perspective of biomarkers discovery, untargeted analysis offers the tantalizing opportunity for novel species discovery and, therefore, the identification of unique, specific, and accurate biomarkers. In untargeted analysis, coverage of the metabolome is unbiased and only restricted by the methodologies of sample preparation and the inherent sensitivity and selectivity of the analytical techniques employed. However, the untargeted approach does suffer from a number of limitations: the protocols and time constraints required to process the extensive raw data sets generated, the analytical chemistry challenges of identifying and characterizing unknown metabolites, the time demands for identifying unknown metabolites that can run weeks and months, the overreliance on the innate analytical coverage of the profiling technique utilized, and a strong bias toward detection of highly abundant small molecule species can all have detrimental effects.

Targeted metabolomics, on the other hand, is supported on the measurement of distinct sets of characterized and chemically annotated small molecules within a biological sample. Absolute quantification of metabolite levels using mass spectrometry can be performed via the standard addition method. Internal standards, typically isotopically labeled versions of the endogenous metabolites, are spiked into the matrix across a range of concentrations to generate a standard curve. The labeled standards, which often contain ^{13}C or ^{2}H isotopes, can be easily differentiated from the endogenous metabolite of interest due to the isotopic difference in mass. It is also imperative that the isotope-labeled standard is pure to avoid interference with, and modulation of, the endogenous metabolite signal. The targeted approach exploits the extensive *a priori* knowledge of a vast array of metabolites and the metabolic pathways to which they belong in order to select specific species for analysis. By analyzing specific key metabolites across all characterized metabolic pathways, these metabolites can function as sentinels to identify perturbations in the pathways, highlighting them for further analysis. When performing targeted metabolomics on a mass spectrometer, the instrument is more sensitive since it is measuring fewer compounds. Furthermore, the sample preparation step can be modified to

reduce the overrepresentation of abundant molecules in the analyses. From a biomarker perspective, this can facilitate the association of low-abundance metabolites with disease states. While their unambiguous identification is known *a priori*, their association with a given phenotype may be novel. If a particular pathway appears to be highlighted in the analyses, additional metabolites in the pathway can be added to the mass spectrometric metabolomics assay for a more detailed interrogation. In addition, since all analyzed species are clearly defined, analytical artifacts are not carried through to downstream analyses, and laborious chemical classification steps are not required.

4. Analytical tools in metabolomics

Regarding the general workflow of an analytical process (Figure 3) that starts from sampling and goes to sample preparation, to detection and, finally, to data treatment; different analytical tools can be used in each step of the process. As data treatment tools are described in the next point, this section is mainly focused on the analytical tools employed in the sample preparation and detection steps.

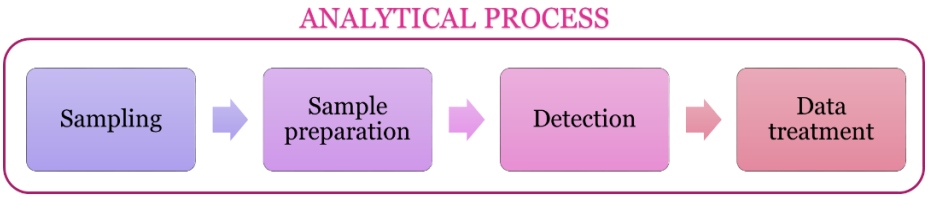


Figure 3. General workflow of an analytical process.

4.1. Analytical tools for sample preparation

Taking into account that sample preparation comprises in most of the cases extraction of the metabolites into a suitable solvent, preconcentration, clean up and, in some cases, derivatization, different methodologies for sample preparation can be used in this step of the analytical process.

Metabolites extraction is a key step in the metabolomics analytical process and its effectiveness directly affects the quality of the final data. The required selectivity of the extraction step will depend on the aim of the study. Therefore, while targeted metabolomics calls for a highly selective extraction that provides clean and concentrated extracts, extraction is eminently non-selective in metabolic profiling; so, only salts and macromolecules are considered potential interferents to be removed in this step. The extraction protocol is mainly conditioned by the target biological sample, which, in the case of solids, is performed by solid–liquid extraction. On the other hand, metabolites from liquid samples are mainly extracted by liquid–liquid extraction (LLE), SPE or solid-phase microextraction (SPME).

The necessity for additional sample preparation steps is marked by the analytical platform that follows this procedure. As an example, the low volatility of many metabolites makes necessary derivatization prior to GC–MS analyses, which is usually performed by silylation [12].

The use of chromatographic techniques prior to detection is of great importance in metabolomics. Despite traditional GC, CE and LC separations have so far been extensively used in metabolomics, new chromatographic approaches have also been implemented with promising perspectives in metabolomics. This is the case of new chromatographic phases like monolithic capillary columns or phases with smaller particle size and higher pressures to increase resolution and sensitivity [18].

4.2. Analytical tools employed in the detection step

Concerning the detection step, as the large number of molecules that constitute the metabolome display a high diversity of chemical structures and abundances, it is impossible to cover the whole metabolome using a single analytical platform. The selection of the detection technique most adequate for each situation is directly dependent on the final purpose of the experiment. Concretely, in targeted metabolomics, where very selective sample treatment procedures must be employed, optical detection techniques such as UV/visible

absorption —using conventional or diode array detectors— or electrochemical techniques such as voltammetry, potentiometry or conductimetry can be used, as well as more selective alternatives such as MS, fluorescence spectrometry or Fourier transform infrared (FT-IR). On the other hand, the most recommended techniques for untargeted metabolomics are NMR spectroscopy and MS due to their sensitivity and resolution capabilities. The latter technique usually requires to be hyphenated to another technique such as GC or LC for separation of the metabolic components [19,20]. Whilst NMR spectroscopy is particularly appropriate for the analysis of bulk metabolites and GC–MS for the analysis of volatile organic compounds and derivatised primary metabolites, LC–MS is highly applicable to the analysis of a wide range of semi-polar compounds including many secondary metabolites of interest.

Benefits and downsides of NMR and MS with respect to their use in metabolomics are depicted in Table 1. These strengths and weaknesses justified that both techniques are needed for metabolome analysis. However, NMR is more suitable for fingerprinting analysis, being MS suitable for all targeted and untargeted metabolomics (profiling and fingerprinting).

As can be seen in the table, one of the strongest points of MS as compared to NMR is its sensitivity. In fact, mass spectrometry-based metabolomics offers quantitative analyses with high selectivity and sensitivity and the potential to identify metabolites with high mass accuracy, taking into account that sensitivity in MS is affected by both ionization and type of detector but, in general terms, it is much more sensitive than NMR.

On the other hand, quantification requires suitable internal standards with similar ionization and fragmentation efficiencies. Concretely, quantitative information on a metabolite peak can be obtained in one of three ways: (i) by integrating it against a reference sample of the same compound (this requires the identity of the metabolite to be known and compared between separate runs, which can introduce error); (ii) by comparing the relative ratios of a set of peaks across a series of spectra; and, (iii) by addition to the sample of a stable-isotope

version of the metabolite of interest. Concerning qualitative analysis, tandem MS fragmentation patterns can give clues as to the connectivity and the presence of specific functional groups in an unknown metabolite. Identification of metabolites is often expedited by comparison to internal standards and searching against mass spectral libraries. Fragmentation of compounds under a certain applied voltage is very reproducible, thus MS/MS measurements are highly selective.

Although MS measurements are fast, the overall run time depends on the chromatographic step, that may vary from minutes to hours (but never to days, as in 2D-NMR). However, combination with chromatographic or electrophoretic equipment reduces the complexity of mass spectra due to metabolite separation in a time dimension, besides delivering additional information on the physicochemical properties of metabolites.

Among the weakest points, mass spectrometry-based techniques usually require sample preparation, which can cause metabolite losses. The complexity of sample preparation depends on the type of mass spectrometer but, in general, it is more tedious than in NMR. However, sample preparation can be automatized in the case of MS.

On the other hand, global profiling may be limited by the ionization type, due to the fact that specific metabolite classes may be discriminated based on the sample introduction system and the ionization efficiency. Despite of this fact, mass spectrometry is currently the most reliable platform for identification as there are a wide variety of libraries with MS and MS/MS information for known components of biological samples.

All those strengths of mass spectrometry against NMR can be verified as mass spectrometry-based metabolomics is characterized by a number of annual publications exceeding those based on NMR.

Table 1. Strengths and weaknesses of NMR and MS in metabolomics.

Nuclear Magnetic Resonance Spectroscopy		Mass Spectrometry	
STRENGTHS	***WEAKNESSES***	***STRENGTHS***	***WEAKNESSES***
✓ Rapid and simple sample preparation protocols.		✓ Possibility of coupling to sample preparation automated systems.	× Laborious sample preparation protocols needed in some cases.
	× Difficult automatization despite robotized systems.	✓ Easy automatization	
✓ No requirement for chromatographic separation.	× Relatively poor resolution in complex samples.	✓ Excellent resolution (improved by separation).	
✓ Useful for identification purposes.	× Difficult to unequivocally assign signals.	✓ Accurate *m/z* measurements ideal for identification of metabolites. ✓ Structural information throughout fragmentation patterns (MS^n).	
✓ High reproducibility.	× Poor sensitivity and selectivity.	✓ Highly sensitive in its different modes. ✓ Very selective (especially in MS^n).	× Detection depends on ionization efficiency.
✓ Non-destructive.			× Destructive, particularly complicated for analysis of valuable samples.
✓ Versatility. Possibility of direct analysis of solids and *in-vivo* studies.		✓ Low sample volume (0.5 to 500 µL).	× Limited to liquid samples or extracts isolated from solid samples.
	× Useful for fingerprinting, less suitable for global profiling and targeted analyses.	✓ Suitable for quantitative targeted analysis, global profiling and fingerprinting.	× Quantification requires chemically-related internal standard.
✓ Relatively high throughput in 1D-NMR experiments.	× Very low throughput in 2D-NMR experiments.	✓ High throughput.	× More efficient when coupled to separation techniques.

4.3. Mass spectrometry systems employed in metabolomics

Mass spectrometry is a detection technique based on the differential displacement of ionized molecules through vacuum by applying an electrical field. Simplistically, a mass spectrometer consists of an ion source, a mass analyzer, a detector and a data system. Sample molecules are inserted to the ion source, where they become ionized. The ions, which are in the gas phase, are separated according to their mass-to-charge ratio (*m/z*) in the mass analyzer and finally detected.

There are different types of ionization sources to generate the gas-phase ions: electron impact ionization and chemical ionization, commonly used in GC–MS, and electrospray ionization and atmospheric pressure chemical ionization, frequently employed in LC–MS. For the purpose of this research, electrospray ionization analyzers have been used when coupling LC to MS, and electron impact ionization in the GC to MS coupling. Electrospray ionization (ESI) is particularly useful to be coupled to LC as ionization occurs at atmospheric pressure. Polar and ionic compounds are best suited for this type of ionization. In ESI, the chromatographic eluate is sprayed from a metal or fused silica capillary. An electrospray is achieved by raising the potential on the spray capillary to 4 kV in positive or negative ionization modes. The resulting spray of charged droplets is directed toward a counter-electrode at a lower electrical potential, where the droplets lose solvent leading to ionic species into the gas phase. The counter-electrode contains an orifice through which ions are transmitted into the vacuum chamber of the mass spectrometer, traversing differentially pumped regions via skimmer lenses [21].

On the other hand, electron impact ionization (EI) is produced by the interaction between energetic electrons and molecules in gas phase to induce ionization and fragmentation of the molecule, producing radical cations.

Once the sample has been ionized, it is transported to the mass analyzer via an electric or magnetic field. The most relevant to the research developed in this Thesis are here briefly described:

i) Triple quadrupole mass spectrometer (QqQ)

A diagram of the design of a triple quadrupole MS is shown in Figure 4 [22].

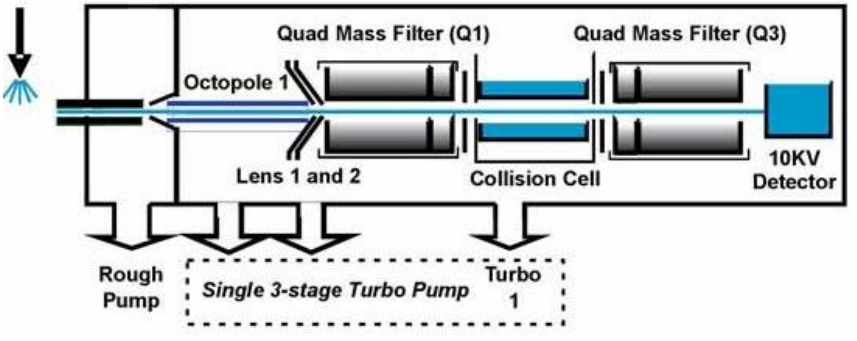


Figure 4. General scheme of a triple quadrupole mass spectrometer.

The triple quadrupole mass spectrometer consists of an ion source followed by ion optics that transfer the ions to the first quadrupole —this device is formed by four parallel rods to which specific direct current and radio frequency voltages are applied. The rods filter out all ions except those of one or more particular *m/z* values as determined by the voltages applied. The applied voltage is variable, so that ions with other *m/z* values are allowed to pass through. Afterwards, selected ions reach a collision cell where they are fragmented. The collision cell is typically called the second quadrupole, but it is actually a hexapole, filled with an inert gas such as nitrogen or argon. The fragment ions formed in the collision cell are then sent to the third quadrupole for a second filtering stage to enable a user to isolate and examine multiple precursor to product ion transitions. This is called *selected reaction monitoring* mode *(SRM)*. Since the fragment ions are pieces of the precursor, they represent portions of the overall structure of the precursor molecule.

Due to the low-mass accuracy achieved with respect to other mass analyzers, triple quadrupole spectrometers are preferably used for targeted analyses, as they allow quantification with high sensitivity and selectivity in SRM mode, actually the most sensitive operational mode for the triple quadrupole MS instrument.

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

Figure 5 shows the diagram of the instrument used in the research in this Thesis [23].

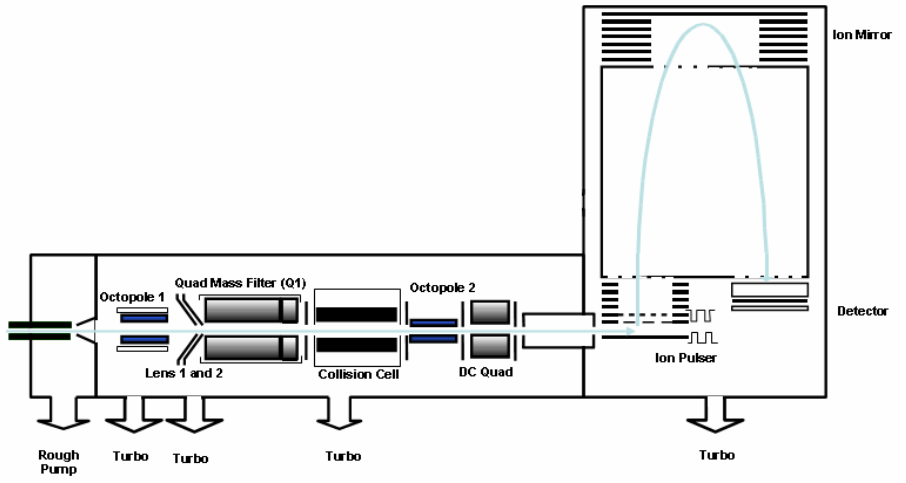


Figure 5. General scheme of a quadrupole time-of-flight mass spectrometer.

The QTOF mass analyzer is based on the same configuration as the QqQ but replacing the last quadrupole by an acceleration tube as mass analyzer (usually in orthogonal configuration) to filter out ions according to the equation of kinetic energy. The QTOF can operate in MS mode with the TOF as scanning tool by taking benefit from the high mass accuracy or in MS/MS mode for structural elucidation.

This hybrid mass analyzer offers better selectivity than triple quadrupoles, meanwhile sensitivity is considerably lower. On the other hand, thanks to its great mass accuracy (below 2 ppm) highly reliable identification can be achieved, thus allowing its use for global metabolic profiling.

iii) Ion trap (IT)

The ion trap is the three dimensional analogue of the linear quadrupole mass filter. In this device, ions are subjected to electrical and magnetical forces

not only in two dimensions as in the simple quadrupole, but also in the third dimension, trapping ions. This feature gives the ion trap the possibility of tandem mass spectrometry experiments and even of performing multiple stage mass spectrometry (MS^n).

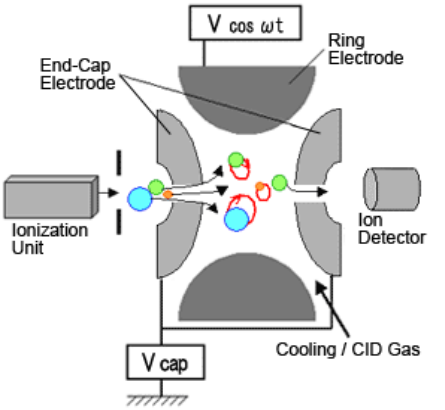


Figure 6. General scheme of an ion trap mass spectrometer.

Figure 6 shows the diagram of the ion trap device [24]. It is composed by a ring electrode and two end-cap electrodes. As can be seen in the diagram, the ions trapped describe orbits inside the trap. The voltage applied in the end-caps (V cap) enables fragmentation of the ions by collisions with the helium damping gas rather than ejection. So, this voltage is used to induce resonance excitation and resonance ejection. On the other hand, the potential applied to the ring electrode determines the range of *m/z* values that can be trapped and the modification of this voltage produces the ejection of the ions for their detection.

5. Data analysis in metabolomics

Statistical considerations play a vital role in the whole process of a metabolomics experiment affecting upon data output quality, analysis and subsequent biological interpretation. Statistical analysis strategies employed in metabolomics are designed according to the approach selected, being slightly different for targeted and untargeted analysis (Figure 7).

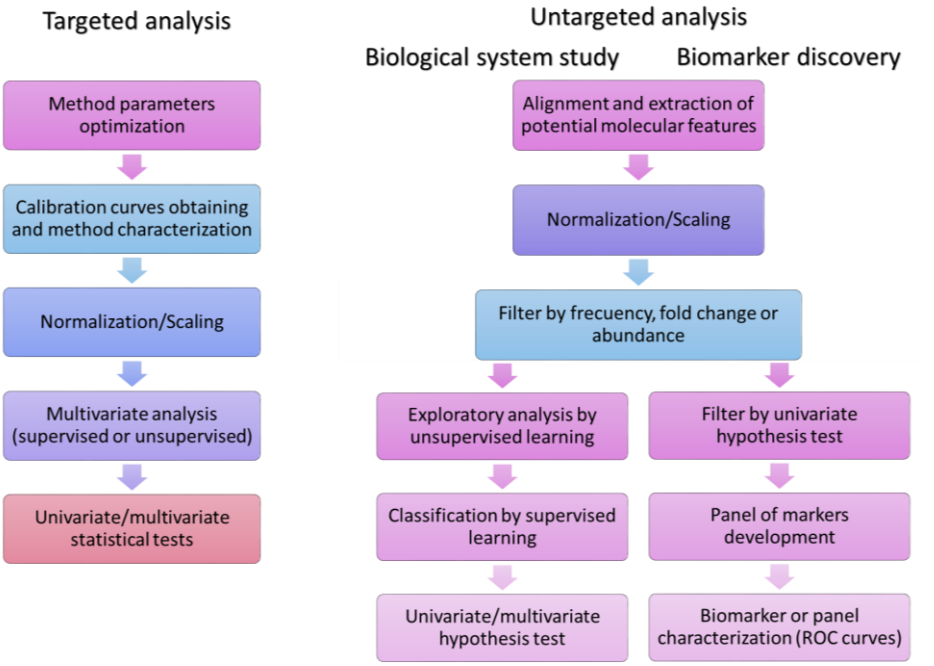


Figure 7. General statistical analysis workflow for targeted and untargeted analysis in metabolomics.

5.1. Statistical analysis for targeted metabolomics

Targeted metabolomics analysis often requires statistical methods to be used during the development of the optimized analytical methodology for the determination of the compounds of interest. Thus, statistical procedures as response surface or screening design are employed to optimize experimental parameters to obtain the best conditions for the determination of the target compounds [25–27]. The rest of statistical analyses employed in the process of developing a new methodology for targeted metabolomics are based on simple statisticals such as simple regression, employed to obtain the calibration curves for metabolites quantitation, or simple calculations, used to characterize the analytical method (reproducibility and repeatability, mainly).

After method application and data generation, common pre-processing strategies as normalization or scaling can be used before interpretation of the results or further statistical studies. These are:

- Normalization: the main objective of this approach is to remove the unwanted systematic bias in ion intensities between measurements, while retaining biological variation. Different normalization techniques can be applied, from normalization using a single metabolite level or standard to logarithmic transformation and more complex alternatives such as probabilistic quotient normalization and quantile normalization [28].
- Scaling: the scaling methods are data pre-treatment approaches that divide each variable by a different factor in order to treat all variables equally, regardless of their intensity. One of the most common factors employed is the mean of each variable, although Z-transform can also be used in case of combining data obtained from different platforms or methodologies [29].

5.2. Statistical analysis for untargeted metabolomics

This kind of metabolomics analysis is focused on covering the maximum number of metabolites using one or more analytical platforms. Thus, generic experimental conditions are mostly employed for sample analysis that is usually done by a separation technique prior to detection. Because of that, after data acquisition, two steps are required: alignment of the chromatograms and extraction of the potential metabolites (usually called potential molecular features) [30]. Owing to the complexity of steps in which many factors need to be considered (adducts, peak shape, noise removal, etc.), different bioinformatics tools have been designed for this purpose.

This type of metabolomics experiments generates large multivariate data sets that constitute a great challenge from a statistical point of view. Thus,

reliable and robust approaches to handle and extract the relevant information from the vast amount of data generated are needed. The outputs of a metabolomics data analysis may differ greatly depending on the purpose of the research. Current metabolomics untargeted studies can be placed into two general categories —those that aim at developing and discovering biomarkers and those that aim at understanding biological processes [31]. In the former case, very few highly dependable metabolites can be sufficient for diagnostic purposes; while in the second case, an extended set of compounds may be desirable when a biochemical network is under examination. Despite the differences between the two strategies, there are also some similarities between them, as in both cases the huge amount of available data requires some filters to facilitate further statistical analysis by reducing the number of potential molecular features. Thus, only potential entities with high intensity, high significance with respect to a factor of interest, or entities that appear with high frequency in the samples batch would pass the filter. As a result, different filters have been created to fulfill the different requirements:

- Fold change analysis: this tool compares the level of each potential molecular feature among the groups under study and retains only the entities that show a minimum change.
- Filter based on univariate hypothesis test: only entities with a *p*-value below 0.05 or 0.01 pass this filter.
- Filter by Volcano plot: this approach is a combination of analysis of variance and fold change analysis.
- Filter by abundance: this tool allows removing all potential molecular features below a certain abundance.
- Filter by frequency: the application of this filter eliminates entities only present below a certain percentage of samples from each group.

i) Untargeted analysis to understand biological processes

Different strategies can be used to study the influence of a known factor (disease, diet, time, genotype, etc.) in a batch of samples and identify the most affected molecular features.

Common chemometric tools [32,33] such as principal component analysis (PCA) [34] are generally used for display and exploratory analysis purposes, whereas univariate statistical tests such as the Student's *t*-test are used to identify the relevant variables after exploratory analysis.

- Exploratory analysis by unsupervised learning: unsupervised methods attempt to analyze a set of observations without measuring or observing any related outcome. As there is not specified class label or response, the data set is considered as a collection of analogous objects. Unsupervised learning uses procedures that attempt to find the natural partitions of 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. The success of such approaches is frequently subjectively evaluated by the interpretability and usefulness of the results with respect to a given problem. PCA is the most common unsupervised method, to the detriment of others such as hierarchical cluster analysis. PCA is an orthogonal transformation of multivariate data first formulated by Pearson [35] mostly used for exploratory analyses by extracting and displaying systematic variations. PCA attempts to uncover hidden internal structures by building principal components describing the maximal variance of the data [36]. This method represents a very useful tool for display 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. The basic concept relies on areas of signal

variance in the data where underlying phenomena can be observed. This principle leads to a focus on a small number of uncorrelated independent signals that explain a large part of the total variance in a compact and insightful manner. In practice, PCA builds hyperplanes in the original space that are linear combinations of the original variables and describes the data in a least squares view. The inspection of PCA scores and loadings plots highlights the relationships among the distribution of samples that may reveal grouping, trends or outliers and the corresponding variables. Moreover, more effective data analyses can be performed on the reduced dimensional space, such as clustering, pattern recognition or classification. The vast majority of metabolomics studies involves PCA as a first exploratory step [37–43].

- Classification by supervised multivariate analysis: supervised learning considers each object with respect to an observed response and includes regression and classification problems depending on the output type under consideration (*i.e.,* a numerical value in the first case and a class label in the second). This classification aims at producing general hypotheses based on a training set of examples that are described by several variables and identified by known labels corresponding to the existing classes. The task is to learn the mapping from the first to the last. Numerous techniques, based either on statistics or on artificial intelligence, have been developed for that purpose. Within the variety of supervised statistical techniques, which includes decision trees, artificial neural networks and support vector machines, the most employed is partial least squares (PLS). This technique is particularly adapted to situations where less observations (N) than measured variables (*e.g.* detected features, K) are available. Its use has become very popular thanks to its ability to deal with many correlated and noisy variables forming megavariate data structures (K>>N) [44, 45]. PLS builds a low dimensional sub-space based on linear combinations of the original X-variables and makes use of the additional Y information by adjusting the

model to capture the Y-related variation in X. A PLS-based classification therefore has the property that it builds data structures with an intrinsic prediction power, by maximizing the covariance between the data and the class assignment. 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, in such as way that a maximum separation among classes is obtained. The model can then be analyzed to understand which variables carry the class-separating information [46]. PLS-DA was demonstrated to be a potent tool for classification of metabolomics data [47].

- Univariate hypothesis testing: the Student's *t*-test, the one-way analysis of variance or the non-parametric equivalents can be used to identify statistical differences between samples of distinct classes [48]. 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. False positives (type I error) are likely to occur when performing multiple comparisons. Procedures such as the Bonferroni of the Benjamini correction have been introduced to address this issue [49]. The vast majority of metabolomics-dedicated software provides statistical hypothesis testing [50,51].

ii) Untargeted analysis to discover biomarkers

The statistical analyses described above are not sufficient to acquire detailed biological understanding, needed for biomarker discovery. In contrast to metabolomics studies focused on deciphering biological processes, where interesting metabolites are found *post hoc*, in biomarker studies metabolite selection should be performed *a priori* rather that *post hoc*. That is, biomarker

selection must be performed before deriving a definitive multivariate predictive model. Furthermore, whereas long lists of metabolites or large multivariate models are quite useful for understanding pathways and biological processes, they are not ideal for developing cost-effective biomarker tests. Rather, a short list of 1–10 biomarkers is mathematically much more robust and far more practical for clinical testing purposes. While pattern discovery methods such as unsupervised clustering or PCA are useful for discovering novel biological processes, they are not ideal for biomarker discovery. Instead, supervised machine learning algorithms (PLS), or multivariate regression models are used for modelling the discriminatory relationship between a binary dependent variable and one or more explanatory variables.

Regarding the selection of potential biomarkers, performing it by univariate statistical significance test is equally inappropriate, as often metabolites that are not significant in isolation can, when combined into a single multivariate model, produce clear and reproducible discrimination. For this reason, iterative methodologies for biomarker panel development, like the employed in the PanelomiX bioinformatics tool or in the ROCCET on line tool, are recommended.

On the other hand, biomarker panels are designed to discriminate with an optimal sensitivity/specificity but in most of the cases the parameter employed for measuring the clinical utility of a panel developed by PLS, for example, is the R^2 (multiple correlation coefficient). Despite R^2 of the model and R^2 of its cross-validation can certainly be used as part of the biomarker selection process, they provide very little transparency and are not a readily interpretable indication of the clinical utility of a given model. Additionally, most clinicians are not familiar with this style of model evaluation, being the ROC (receiver operating characteristic) curve analysis the standard method for describing and assessing the performance of medical diagnostic tests with binary classification [52].

ROC curves consider the frequency with which the test produces true positives, true negatives, false positives and false negatives. These values are

summarized into the proportion of actual positives that are correctly classified as positive (sensitivity) and the proportion of actual negatives that are correctly classified as negative (specificity). An example of ROC curve is shown in Figure 8.

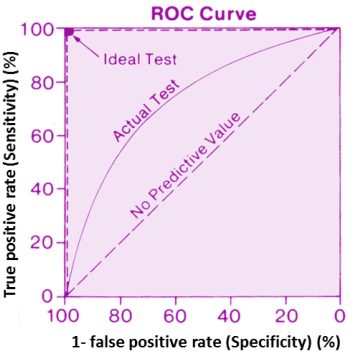


Figure 8. ROC curve representation including the curve described for an ideal test, and a test without predictive value.

For ease of interpretation, 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 false positives (the negative result of the test from a subject that has an actual negative outcome). 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 affect sensitivity and specificity is through a ROC curve. Furthermore, the ROC curve is a non-parametric measure of biomarker utility, being widely considered the most objective and statistically valid method for biomarker performance evaluation [52,53].

There are different informatics tools for ROC curve generation and the parameter that often summarized the curve is the area under the curve (AUC). This parameter can be interpreted as the probability that a diagnostic test or a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. If all positive samples are ranked before negatives ones (example of a perfect classifier), the AUC is 1.0. An AUC of 0.5 is equivalent to randomly classifying subjects as either positive or negative (classifier without

utility). However, using the whole area under a ROC curve may not be appropriate in some cases and the partial AUC is most useful when only certain regions of the ROC space (high sensitivity or high specificity) are of particular interest [54].

6. Present role of metabolomics in clinical research: personalized medicine

Despite metabolomics is the most recent of the primary omics, its significant contribution to the clinical field deserves to be discussed as well as its key role in the incipient personalized medicine (PM). The secondary omics in which the most salient contributions have been developed are lipidomics and nutrimetabolomics.

In the omics world, lipids have long been in the shadows, while nucleic acids and proteins hogged the limelight. Recent improvements in LC and MS, together with the recognized role of lipids functions in most of the so-called "diseases of the industrialized societies", have provided the rapid expansion of lipid research. Lipids fulfill a variety of vital functions, as they are central constituents of biological membranes, serve as energy storage compounds, and act as second messengers or as covalent modifiers governing the localization of proteins and favoring effectiveness of protein–protein and lipid–protein interactions. In addition, bioactive lipids generated during remodeling of membrane lipids by activated lipases serve as intra and extracellular mediators in cell signalling, in cell–cell communication, inflammation, host-defence mechanisms and ischemia-reperfusion, apoptosis and response to stimulus [55].

Lipidomics, defined as a comprehensive analysis of all lipids in a biological system (lipidome), is, through lipid profiling, a fast-growing area of research due to the involvements of lipids in human diseases such as obesity, atherosclerosis, high-pressure and diabetes (known as "metabolic syndrome"), Alzheimer's disease, and as well as in drug discovery [56–62].

Nutrimetabolomics or nutritional metabolomics is one of the most firm supports of the intended PM. The close relationship between nutrition and health —each time more defended by clinicians— is clearly demonstrated by metabolomics studies.

Particularly important in this field is the role of metabolomics to decipher gut microbial metabolic influence on health and disease, as the consortium of symbiotic gut microorganisms can be viewed as a metabolically adaptable, rapidly renewable, and metabolically flexible ecosystem varying in addition with the host's age, diet, and health status. Because metabolites are the products and by-products of the many intricate biosynthetic and catabolic pathways existing in all living systems, monitoring the resulting metabolic variations provides a unique insight into intra- and extra-cellular regulatory processes involved in our metabolic homeostasis. By the study of low molecular weight compounds in biofluids (blood and urine), stools, and tissues (intact biopsies or extracts), metabolomics assures the characterization of metabolic fingerprints that can be associated with individual phenotypes, which encompass dietary or disease status.

The intestinal tract is one of the most important interfaces between mammalian metabolism and the environment along which the varying metabolic activities of the gut microbiota not only determine absorption, digestion, metabolism, and excretion of dietary nutrients but also shape regio-specifically the surrounding and distant host cell biochemical processes [63]. Advances in metabolomics applications are providing novel insights into the molecular foundations of these host–microbial relationships and their influence onto health and disease risks [64, 65]. In particular, a series of investigations in humans [66], rats [67], and gnotobiotic mice [68] have provided a set of reference metabolic profiles of gut intestinal biopsies that can be used not only to assess compartment structure and function but also the gut microbial impact at the tissue level [69]. Therefore, these studies indicated that the type of gut microbiota may be a key factor in the determination of the intestinal homeostasis, osmo-protection, motility, and calorie recovery from the diet.

Food products are nowadays more and more commonly employed to modulate the composition of the gut microbiota with the main objective to prevent disease development and contribute to the health status' maintenance. The effects of consuming live microbial supplements (probiotics) on the microbial ecology, and on human health and nutritional status have been extensively investigated over the past years [70]. Probiotic supplementation aims at replacing or reducing the number of potentially pathogenic bacteria in the intestine by enriching the populations of beneficial strains that ferment carbohydrates and have little proteolytic activity. As an alternative, prebiotics, non-digestible food ingredients, generally oligosaccharides, modify the balance of the intestinal microbiota by stimulating the activity of health beneficial bacteria, such as lactobacilli and bifidobacteria. Eventually, the combined use of prebiotics and probiotics, also called "synbiotics", may offer superior effects in health maintenance through modulating the microbial functional ecology [70].

The influence of the gut microorganisms on the progression of human diseases is nowadays a main concern in the etiology and/or maintenance of gut dysfunctions, such as irritable bowel syndrome (IBS) [71] or inflammatory bowel disease (IBD) [72]. The incidence and prevalence of these diseases make mandatory more research on the gut–metabolomics binomial [73].

Metabolomics was also proven to be a valuable diagnostic tool to differentiate Crohn's disease (CD) from ulcerative colitis (UC), but also active and quiescent UC, as per the analysis of intact gut biopsies and colonocytes [74].

A recent study associates consumption of black tea with reduced cardiovascular risk. Black tea polyphenols (BTPs), high-molecular-weight species that predominantly persist in the colon, can undergo a wide range of bioconversions by the resident colonic microbiota and can in turn also modulate gut microbial diversity. Novel metabolomics platforms, coupled to *de novo* identification currently available, have been used to cover the large diversity of BTP bioconversions by the gut microbiota. Evidence for cardiovascular benefits of BTPs points toward anti-inflammatory and blood pressure-lowering properties

and improvement in platelet and endothelial function for specific microbial bioconversion products. It has also been of interest to assess how phenotypic variation in gut microbial BTP bioconversion capacity relates to gut and cardiovascular health predisposition [75].

A salient comment deserves an in development study that, under the acronym NU–AGE (from "new dietary strategies addressing the specific needs of elderly population for a healthy aging in Europe) involved a number of European researchers. The NU-AGE rationale is that a one-year Mediterranean whole diet (considered by UNESCO a heritage of humanity), newly designed to meet the nutritional needs of the elderly, will reduce "inflammaging" in about 3.000 fully characterized subjects aged 65–79 years of age, and will have systemic beneficial effects on health status (physical and cognitive). Within the comprehensive set of analyses that will be performed to identify the underpinning molecular mechanisms metabolomics profiling and targeted analyses play a key role [76].

Clinical metabolomics has different purposes, including disease diagnosis and follow-up, therapeutic drug monitoring and PM development. Within the different applications for disease diagnosis, the most common medical areas in metabolomics are cardiology, human reproduction, diabetes, central nervous system diseases and oncology. Studies for disease diagnosis involve mainly the discovery of new biomarkers and their characterization. The search for and use of biomarkers in metabolomics is described in deep in the next section. However, there are also metabolomics clinical studies, the objective of which is to understand or find which metabolic pathways are perturbed by the presence of a given disease. This research has allowed to know, as an example, how lipid metabolism is highly perturbed by Alzheimer's disease [77].

As mentioned before, diabetes and cardiovascular diseases are two of the most studied diseases by metabolomics. The worldwide pandemic rise of obesity prevalence is strongly linked to incidence of type2 diabetes (T2D) and cardiovascular disorders. Recently, a branched chain amino acids (BCAAs) related metabolic signature contributing to insulin resistance in obese human

subjects has been described [78]. Furthermore, metabolomic analyses have been applied to the study of an epidemiological cohort in which diabetes-related complications could be detected already under sub-clinical conditions [79]. In addition to previously reported T2D biomarkers, including sugar metabolites, ketone bodies, and BCAA, metabolites resulting from perturbations of metabolic pathways linked to kidney dysfunction (3-indoxyl sulfate), lipid metabolism (glycerophospholipids, free fatty acids), and bile acid metabolism have been considered.

Moreover, metabolomics was employed to decipher indicators of early onsets of pre-diabetes status, marked by alterations in fatty acid, tryptophan, uric acid, bile acid, and gut microbial metabolism [80]. This information can be easily obtained by using the sampling in the oral glucose tolerance test (OGTT) that, as it is well known, only provides the evolution of glucose and insuline along the test. In addition to the evolution of these parameters, monitoring the metabolic evolution of the OGTT provides information on the lactate, hyppurate and glycerol, and, more interesting, on the changes of levels of three bile acids — glycocholic acid, glycochenodeoxycholic acid, and taurochenodeoxycholic acid— which provides information on key parameters related to cholesterol levels [81, 82]. Increases in lysophosphatidylcholine [80] and decreases in amino acids [83], acylcarnitines [80] and fatty acids [80,83] have also been reported. Although the levels of fatty acids declined during an OGTT, the levels of saturated (SFA) and monounsaturated fatty acids (MUFA) were more significantly decreased than those of polyunsaturated fatty acids (PUFA). Moreover, a substantial reduction in the SFA/MUFA ratio was observed, consisting of a shift from MUFA towards SFA. These findings indicate a change in fatty acid composition following an OGTT [83].

Metabolomics studied related to respiratory disorders have been increased in the last few years through the study of organic volatile compounds contained in the exhaled breath. As an example, comparison of exhaled air profiles from individuals with impaired respiratory function revealed differences between participants with and without chronic obstructive pulmonary disease

[84]. However, the identification of the compounds responsible of these differences has not been done.

In dealing with PM, it may be said that the current revolution in metabolomics will at term offers new opportunities for preventive medicine, prognostic strategies and provides a new step towards PM. The importance of any branch of metabolomics in the search for PM is demonstrated by the presence of this omics in all present definitions of PM [85]:

- Personalized Medicine seeks to improve tailoring and timing of preventive and therapeutic measures by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.
- Personalized Medicine seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics.

Some recent contributions also deserve to be commented just to show the wide variety of clinical (or close to clinical) aspects where metabolomics can be a useful tool. Complementary and alternative medicine has also taken advantage from metabolomics. As an example, aromatherapy that uses essential oils through inhalation had not evidenced its efficacy in treating medical conditions owing to a particular lack of studies employing rigorous analytical methods that capture its identifiable impact on human biology. Recently, a comprehensive metabolomics study has revealed changes in people, after exposed to aroma inhalation for 10 continuous days, consisting of metabolic alterations in urine from people with anxiety symptoms. The samples, analyzed by GC–TOF and ultra performance liquid chromatography (UPLC) coupled to QTOF, showed a significant change of the metabolic profile of the individuals before and after subjected to aromatherapy. The change was characterized by the increased levels of arginine, homocysteine, and betaine, as well as decreased levels of alcohols, carbohydrates, and organic acids in urine. Notably, the metabolites from

tricarboxylic acids cycle and gut microbial metabolism were significantly altered. This study demonstrates that the metabolomics approach can capture the subtle metabolic changes resulting from exposure to essential oils, which may lead to an improved mechanistic understanding of aromatherapy [86].

7. The search for and use of metabolomic biomarkers

7.1. Definition and features of biomarkers

Cells, either directly or indirectly (via extracellular fluid), communicate with body fluids. Cell metabolites, peptides and proteins are released from cells or are taken up from body fluids via normal excretion, trans-membrane diffusion or transport, and throughout the death process during which cells release all of their contents. Thus, at least to a certain extent, biochemical and protein changes in cells and organs are reflected in body fluids. While tissue samples, biopsies, and certain fluids such as urine (kidney), bile (liver), and cerebrospinal fluid mainly reflect changes in specific organs and thus are considered "proximal matrices"; plasma samples reflect systemic changes that often cannot be traced back to a certain organ [87]. Such changes of metabolites, peptides and proteins in body fluids, if mechanistically linked to disease processes and drug effects in tissues and organs, have the potential to serve as surrogate markers or biomarkers.

In 2001, a consensus panel at the National Institute of Health defined the term biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenesis processes, or pharmacologic responses to a therapeutic intervention" [88]. Despite biomarkers are not new in the field of medicine [89], they have gained immense scientific and clinical interest in recent years thanks, in many aspects, to omics disciplines.

Biomarkers are useful along several points of a disease continuum. Thus, they are useful in the context of primary prevention, for preventing disease itself. Also, they can facilitate secondary prevention for the early detection of disease via

screening, detection of subclinical disease and by aiding the monitoring of disease progression. Finally, biomarkers are also useful for the purpose of tertiary prevention, helping guide treatment to avoid morbidity owing to established disease [90].

One avenue for the identification of novel risk markers is being opened by the global analysis of the human metabolome. While decades of research in biochemistry, nutrition, and physiology have revealed specific metabolic pathways, systematic surveys of pathways altered in human disease states are now possible. An emerging set of metabolite profiling capabilities, based on techniques such as MS and NMR spectroscopy, enable the monitoring of hundreds of analytes from biological samples. These technologies promise to transform our ability to profile samples, with the goal of illuminating biology and discovering valuable clinical biomarkers.

Today's clinical diagnostics is typically based on a limited set of biomarkers, often only one parameter that is associated with the functional aspect of an organ or a specific disease process (for example creatinine in serum as a marker of kidney function). However, there are no single molecular entity markers, and there will never be one that captures the function of organs such as the kidney, liver or vascular endothelium in all its complexity.

Biomarkers are useful as research and clinical tools in most major medical fields, as shown in Table 2.

The characteristics of an ideal biomarker are the following:

1) Safe and easy to measure, in such a way that the measurement could be repeated, if required, without danger for the patient and be performed by not very specialized personnel.

2) Low cost to measure, not only that of the test, but also the cost of follow-up testing should be low to monitor future levels of the biomarker and treatment efficacy.

3) The occurrence of a moderate proportion of the disease in the community should be explained by the biomarker.

4) High predictive accuracy for detecting a disease. The sensitivity (the detection of a disease when that disease is truly present) and specificity (recognition of true absence of disease) of the biomarker should be relatively high, thus reducing false-positive and false-negative rates.

5) Consistent across sexes and ethnic groups.

Table 2. Examples of biomarkers currently used in different fields of medicine.

Specialization	**Disease**	**Screening to identify at-risk patients (primary prevention)**	**Diagnostic for existing disease states (secondary prevention)**	**Prognostic for treatment and outcomes (tertiary prevention)**
Cardiovascular	Cardiovascular disease	Serum LDL cholesterol	Coronary artery computed tomography screening	Echocardiographic left ventricular ejection fraction as a prognosticator for sudden cardiac death
Endocrinology	Diabetes mellitus	Fasting blood glucose	Microalbuminuria screening for diabetic nephropathy	Hemoglobin A_{1c}% and long-term glucose control
Gastroenterology	Colon cancer	Family history of premature colon cancer	Colonoscopy	Carcinoembryonic antigen for postsurgical colon cancer recurrence
Rheumatology	Systemic lupus erythematous	-	Anti-Smith antibody specific for detecting lupus	Andi-dsDNA for diagnosis of lupus nephritis
Oncology	Breast cancer	BRCA-1 gene BRCA-2 gene	Mammography	Tumor estrogen receptor, progesterone receptor and HER2-neu receptor status for decisions about tamoxifen therapy

It should be noted that a biomarker may be very useful even if it does not meet all the criteria of an ideal biomarker. For example, high density lipoprotein (HDL) cholesterol protects against cardiovascular disease (CVD); however, raising HDL cholesterol has not been proven to decrease CVD risk. On the other hand, positive and negative predictive values are dependent on the prevalence of disease in a given population: as the prevalence of disease within a given population declines, the positive predictive value gets smaller and the negative predictive values gets larger for the same test. Predictive values of biomarkers with regards to specific outcomes are also important to consider. For instance, troponin is a highly sensitive biomarker of early acute myocardial infarction mortality [91]; therefore, both the diagnostic and predictive aspects of troponin make it an ideal biomarker for detecting this disease.

Even if a biomarker meets several criteria that make it "ideal", this does not imply that the biomarker will necessarily be useful in a clinical setting. Specifically, if a novel biomarker cannot add a value in clinical settings, then it may never pass the sizeable hurdle that separates clinical practice from clinical research.

7.2. The analytical platforms for biomarkers implementation

Multiplexing analytical technologies, including but not limited to, arrays, bead immunoassays and mass spectrometry, allow for the assessment of molecular marker panels ideally in a single run. Among these technologies MS is attractive due to its sensitivity, specificity and flexibility, and it allows for absolute quantification. Whereas antibodies are derived from biological sources, MS-based multiplexing assays lack the manufacturing and batch-to-batch reproducibility challenges of many antibody-based assays. While in antibody-based assays the type and number of compounds detected is always limited by the antibodies included in the assay, mass spectrometry provides the basis for non-targeted approaches and, therefore, the means for detection of an open number of previously undefined signals [92]. Non-targeted assays constitute the extreme of multiplexing assays and, in an ideal world, would capture the

complete human metabolome or proteome. The major difference between an untargeted and targeted assay is that in a truly untargeted analysis, the compounds underlying the signals are not known. Thus, such assays are semi-quantitative at best and can only be validated to a limited extent. In contrast to untargeted assays, targeted multiplexing assays simultaneously measure several well-defined compounds, which are validated and are quantitative. Targeted multiplexing assays avoid some of the analytical and statistical uncertainties associated with completely untargeted data sets in terms of the quality of data, quantification, interferences as well as false negatives and positives.

Both the platform used to obtain the raw analytical data and the chemometric tools used for data treatment, together with the corresponding data bases are keys to success in the search for biomarkers [93].

7.3. The process for biomarker development

Regulatory agencies have established review structures as well as guidelines that outline the biomarker qualification process [94]. Based on these guidelines, the biomarker development process can be divided into three stages: discovery, verification and qualification.

Discovery, as the first stage, can be divided into two sub-stages, as Fig. 9 shows: identification of the candidate biomarker, and analytical validation. The former is an untargeted analysis to identify the potential biomarkers. This procedure is commonly carried out by high resolution mass spectrometry as QTOF. After extraction and identification of the potential markers, analytical validation involves a targeted step in which the predictive behavior of the markers is properly established by using more sensitive methods generally involving the use of QqQ.

Verification mechanistically links the molecular marker to the biochemical process underlying a disease or drug effect. The qualification process bridges the results of molecular marker measurements, symptomatic drug effects and disease outcomes.

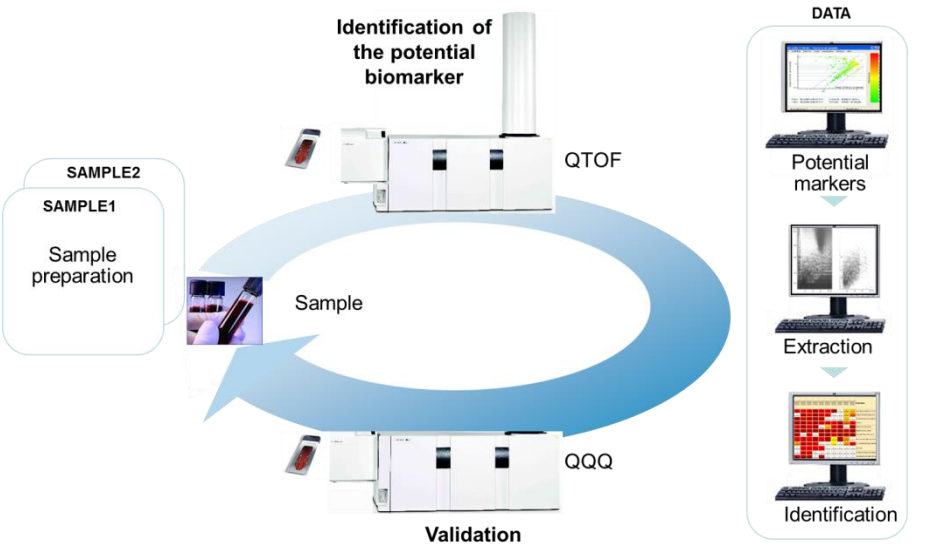


Figure 9. Analytical platform for identification/validation of biomarkers.

Qualification has been defined as "a graded, fit-for purpose evidentiary process linking a biomarker with biology and clinical endpoints" [95]. Qualification has been differentiated from validation, which focuses on the reliability and performance characteristics of the analytical assay used to measure molecular markers [94,96]. Translation of a molecular marker from the discovery stage into pre-clinical testing and clinical development greatly depends on the availability of robust, precise and sensitive assays for the measurement of a larger number of samples [97]. While during the discovery phase a partial validation following "fit-for-purpose" principles will be sufficient, the validation must become more stringent if drug development strategies, regulatory approval and clinical decisions will depend on such a molecular marker. During later clinical development, and especially when developed as a clinical diagnostic tool, which typically involves comparison of individual results with normal values, absolute quantification of the analytes and complete validation following applicable regulatory guidelines is critical. If appropriately qualified and based on adequately validated assays, molecular markers can support primary outcomes, they may help to understand and monitor mechanisms of toxicity, drug

interactions, disease-drug interactions and the effects of genotypes, sex and age. They can be used to stratify patient populations, guide subgroup analyses to bridge safety and efficacy data between different populations such as adults to pediatric patients, and among different ethnic groups. Molecular markers may also be developed into clinical diagnostic tests and, finally, they must comply with a set of rules and regulations for premarketing clearance or premarket approval oversight by the corresponding organism in the application country [98–101].

7.4. Metabolomics biomarkers

Because the vast number of diseases and the number of biomarkers reported for each, two (*viz.,* CVD and cancer) have been the diseases selected to comment metabolomic biomarkers reported either for them or for related risk factors.

Metabolomics has been used to study various forms of *cardiovascular risk factors* such as diabetes mellitus, obesity, and metabolic syndrome [102–104]. There is no doubt that the growing prevalence of obesity has skyrocketed the interest in identifying biomarkers of this disease, as obesity is considered a major risk factor for CVD [105]. A large number of potential obesity biomarkers that correlate with traditional cardiovascular risk factors or predict subsequent cardiovascular events have been reported (usually involving *in vitro* and *in vivo* studies in animals, then in humans), but they require an in-depth validation in humans [106].

The also growing prevalence of type 2 diabetes mellitus (T2DM) is a consequence of obesity and sedentary life habits, which correlate with CVD. Plasma fatty acids metabolic profile coupled to uncorrelated linear discriminant analysis has been reported as biomarker screening of T2DM and type 2 diabetic coronary hearts diseases [107].

Present research on *biomarkers associated with common pathologies in CVD* [108] ranges from markers for myocardial ischemia [109–111] and cardiogenic shock [112], risk of developing atherosclerosis or future cardiovascular events [113–115], chemotherapy-induced cardiotoxicity [116], and

pulmonary hypertension related to advanced heart failure [117]. Nevertheless, at present, only biomarkers for the consequences of ischemia (*i.e.,* myocardial necrosis) are available. Abnormal levels were found for γ-aminobutyric acid, uric acid, citric acid, and several yet to be identified metabolites [111]. To extend beyond ischemia, the plasma signature of aconitic acid, hypoxanthine, trimethylamine-N-oxide, and threonine differentiated with high diagnostic accuracy between infarcted subjects and patients undergoing coronary angiography [110]. Despite the studies on the role of NO in the pathogenesis of cardiogenic shock [118], the metabolomic insights of disregulated NO metabolism links to cardiogenic shock has not been demonstrated. No conclusive metabolomics studies to identify risk in patients with atherosclerosis and its associated adverse events have involved from phosphatidylcholine and choline metabolism to disregulation of the arginine-NO metabolic pathways [114].

On the other hand, *lung cancer* is the leading cause of cancer-related death in most developed countries [119], as a consequence of the fact that 60% of patients are diagnosed at advanced stages when a cure is unlikely [120]. The annual mortality rate for lung cancer exceeds the annual rate for breast, prostate, and colon cancer combined, all of which have successful clinical screening tools for the detection of early-stage disease [121]. Therefore, the search for diagnostic strategies for early lung cancer detection has intensified in the last decade. Early detection involves a high-risk population, a screening test, and a testing schedule.

Within this context, one must distinguish populations of individuals at-risk before or after the disease becomes measurable (Figure 10).

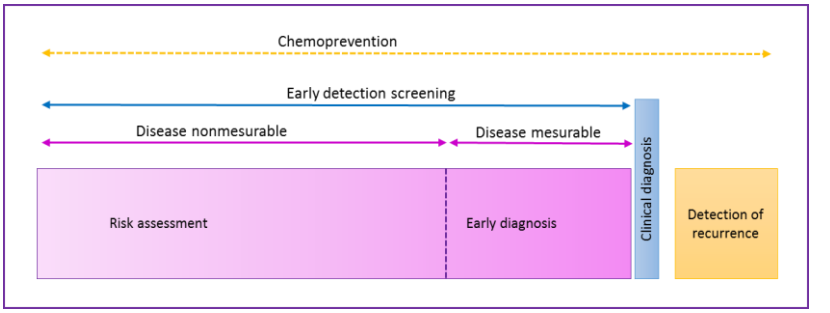


Figure 10. Clinical contexts for biomarker development in early detection of lung cancer.

The search for early detection of lung cancer biomarkers has involved the following candidate samples–omics:

(i) *Tissues*, and genomics —hypermethylation of a number of tumor suppressor genes [122–124)—, transcriptomics —regions of chromosomal amplification [125], mRNA expression variation [126,127], the differential expression of several microRNAs [128]—, proteomics —the proteomic signature of invasive [129] and pre-invasive lesions in lung tissues [130]— and, of course, metabolites, which are considered in more detail.

Metabolomic profiles of lung and prostate tissues, obtained by CE–TOF/MS, comprised 114 and 86 metabolites, respectively, and the profiles not only well distinguished tumor from normal tissues, but also squamous cell carcinoma from the other tumor types in lung cancer and poorly differentiated tumors from moderately differentiated tumors in prostate cancer. Concentrations of most amino acids, especially branched-chain amino acids, were significantly higher in tumor tissues, independent of organ type, but of essential amino acids were particularly higher in poorly differentiated than in moderately differentiated prostate cancers. Organ-dependent differences were prominent at the levels of glycolytic and tricarboxylic acids cycle intermediates and associated energy status. Significantly high lactate concentrations and elevated activating phosphorylation levels of phospho-fructokinase and pyruvate kinase in lung tumors confirmed hyperactive glycolysis [131].

(ii) *Biofluids* including peripheral blood and its components (circulating cells, plasma, and serum), exhaled breath condensate, urine, and sputum offer non-invasive access to large quantities of samples available for analysis. Alterations can lead to the generation of disease-specific molecular species such as altered or methylated DNA, overexpressed mRNA, miRNA, or proteins that can potentially be released into the extracellular microenvironment. Therefore, molecular analyses of early

stage lung cancer-related biofluids represent an attractive choice for the discovery and validation of diagnostic biomarkers [132,133].

Exhaled breath condensate (EBC) is the cooling of exhaled gas to gain insight into the composition of extracellular lining fluid and soluble exhaled gases. Compounds that have so far been measured include, in addition to proteins and DNA [134], lipid peroxidation products, products of nitrogen oxide metabolism, hydrogen ions, hydrogen peroxide, and cytokines. EBC represents the non-invasive sampling for diagnosing lung cancer, which offers a wider potential on metabolomics research. The analysis of volatile organic compounds (VOC) that underlying rationale of this sampling approach is based on the observation that tumor cell growth is accompanied by the alteration of protein expression pattern that may lead to peroxidation of the cell membrane and thus to the emission of VOCs [135]. Several recent studies have used GC–MS analysis of VOCs as both discovery and validation platforms [136–138]. Other groups made use of the analytical power of GC–MS and the sensitivity of custom designed nanosensors in which changes in electrical resistance from organic compounds contained in exhaled breath of patients can be detected by these sensors and recorded. For example, in a study by Peng and colleagues, a VOC signature that distinguished patients with lung, colorectal, and breast cancers from healthy individuals was recently identified from exhaled alveolar breath [139–141]. Other studies attempted to identify volatile proteins and peptides present in EBC and used them as potential markers for the early detection of lung cancer [140, 141]. The results of these studies provide evidence for feasibility of this strategy to isolate and identify proteins useful for early detection of lung cancer. Further studies are still needed to standardize a collection device, to further show specificity of any test, and to determine the use of this approach in clinical practice. Some achievements of metabolomics in the search for lung cancer biomarkers are as follows:

- Decreased concentrations of alkanes in the breath of patients with lung cancer as compared to samples obtained from controls have been detected [142]. The explanation for this finding has been that, during carcinogenesis, cytochrome P450 enzymes (CYP) are induced resulting in the enhanced catabolism of several VOCs. The test for these biomarkers has a sensitivity of 85% and a specificity of 80% in selecting patients.

- Inhibition of proliferation of human lung, colon and pancreatic cancer cells by prostaglandin E3 (PGE3) —derived from cyclooxygenase 2 (COX-2) metabolism of the omega-3 fatty acid eicosapentaenoic acid (EPA)— has also been reported [143]. However, how PGE3 metabolism is regulated in cancer cells, particularly in human non-small cell lung cancer (NSCLC) cells, had not been fully understood, until identification, using MALDI, of differences in lipid metabolism between two human NSCLC cell lines, A549 and H596. This identification could contribute to their differential response to EPA treatment, which showed that the level of EPA incorporated into phospholipids in H596 cells was 4-fold higher than A549 cells. Intriguingly, H596 cells produced much less PGE3 than A549 cells even though the expression of COX-2 was similar in these two cell lines.

- Lipids peroxidation, a well-known index of free radical activity, occurs as a result of auto-oxidation of polyunsaturated fatty acids and it has been implicated in the pathogenesis of lung cancer [144]. Khyshiktyev *et al.* have reported that, without stating the actual marker, levels of lipid peroxidation were lower in EBC of patients with lung cancer than in controls [145].

- Significant differences in EBC hydrogen peroxide levels were demonstrated between groups: lung cancer subjects (23.68±9.15 μM); smokers (5.21±0.69 μM); ex-smokers (14.35±3.79 μM); and non-smokers (17.59±6.53 μM) [146].

- Different samples such as serum (from 29 healthy volunteers and 33 lung cancer patients with adenocarcinoma, n = 12; squamous cell carcinoma, n = 11; or small cell carcinoma, n = 10 ranging from stage I to stage IV disease), and lung tissue (from 7 lung cancer patients including the tumor tissue and its surrounding normal tissue) were analyzed and a total of 58 metabolites (57 individual metabolites) were detected in serum, and 71 metabolites were detected in the lung tissue. The levels of 23 of the 58 serum metabolites were significantly changed in all lung cancer patients compared with healthy volunteers, and the levels of 48 of the 71 metabolites were significantly changed in the tumor tissue compared with the non-tumor tissue. Partial least squares discriminant analysis, applied to the serum sample data, showed characteristic alterations in each histological subtype and disease stage [147].

Special mention deserves the recent use of sweat as sample to search for lung cancer biomarkers, as discussed in Chapters 1 and 7 of this PhD book [148–150].

Finally, two general aspects in dealing with present research on biomarkers deserve to be emphasized:

(i) The general mandate for orthogonal (*i.e.,* uncorrelated) disease biomarkers that provide additional clinical information along new biological axes. A robust set of predictors for identifying at risk individuals is of particular importance because of the delay or prevention of diseases such as atherosclerosis. Therefore, existing biomarkers can be combined with present ones to offer higher sensitivity and/or selectivity and, in short, more accuracy in the resulting biomarker.

(ii) The indubitable certainty that systems biology may provide the best and closest to ideal biomarkers. Attempts in this sense by combining proteomics and metabolomics enable to link alterations of cellular proteins to metabolism and function [151].

References

[1] J.K. Nicholson, I.D. Wilson, Opinion: Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. *Nature Rev. Drug Disc.* 2 (2003) 668–676.

[2] J.K. Nicholson, J.C. Lindon, Systems biology: metabonomics. *Nature* 455 (2008) 1054–1056.

[3] E. Holmes, I.D. Wilson, J.K. Nicholson, Metabolic phenotyping in health and disease. *Cell* 134 (2008) 714–717.

[4] L.C. Heather, X. Wang, J.A. West, J.L. Griffin, A practical guide to metabolomic profiling as a discovery tool for human heart disease. *J. Mol. Cell Cardiol.* 55 (2013) 2–11.

[5] R. Rosewell, C. Vitols, Identifying metabolite in biofluids. Available online: http://www.chenomx.com/news/img/page23/ChenomxAppNote.an001.pdf (accessed on 20 March 2013).

[6] L. Atzori, R. Antonucci, L. Barberini, E. Locci, F. Cesare Marincola, P. Scano, P. Cortesi, R. Agostiniani, A. Weljie, A. Lai, V. Fanos, 1H NMR-based metabolic profiling of urine from children with nephrouropathies. *Front. Biosci.* 2 (2010) 725–732.

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

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

[9] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallow, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0 – The human metabolome database in 2013. *Nucleic Acids Res.* 41 (2013) D801–D807.

[10] W.C. Levy, D. Mozaffarian, D.T. Linker, S.C. Sutradhar, S.D. Anker, A.B. Cropp, I. Anand, A. Maggioni, P. Burton, M.D. Sullivan, B. Pitt, P.A. Poole-Wilson, D.L. Mann, M. Packer, Heart Failure. The Seattle heart failure model. Prediction of survival in heart failure. *Circulation* 113 (2006) 1424–1433.

[11] G.C. Fonarow, K.F. Adams, W.T. Yancy, W.J. Boscardin, Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tress analysis. *JAMA* 293 (2005) 572–580.

[12] W.B. Dunn, D.I. Ellis, Metabolomics: Current analytical platforms and methodologies. *Trends Anal. Chem*. 24 (2005) 285–294.

[13] K.T. Myint, K. Aoshima, S. Tanaka, T. Nakamura, Y. Oda, Quantitative profiling of polar cationic metabolites in human cerebrospinal fluid by reversed-phase nanoliquid chromatography/mass spectrometry. *Anal. Chem*. 81 (2009) 1121–1129.

[14] 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.

[15] E. Allard, D. Bäckström, R. Danielson, P.J.R. Sjöberg, J. Bergquist, Comparing capillary electrophoresis-mass spectrometry fingerprints of urine samples obtained after intake of coffee, tea or water. *Anal. Chem*. 80 (2008) 8946–8955.

[16] D.B. Kell, M. Brown, H.M. Davey, W.B. Dunn, I. Spasic, S.G. Oliver, Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol*. 3 (2005) 557–565.

[17] H.G. Gika, G.A. Theodoridis, I.D. Wilson, Liquid chromatography and ultra-performance liquid chromatography–mass spectrometry finger-printing of human urine: Sample stability under different handling and

storage conditions for metabonomics studies. *J. Chromatogr. A* 1189 (2008) 314–322.

[18] V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal. Chem.* 75 (2003) 6737–6740.

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

[20] F. Dieterle, B. Riefke, G. Schlotterbeck, A. Ross, H. Senn, A. Amberg, NMR and MS methods for metabonomics. *Methods Mol. Biol.* 691 (2011) 385–415.

[21] W.J. Griffiths, Metabolomics, metabonomics and metabolite profiling. *RSC Publishing*, 2007.

[22] J. Ni, J. Rowe, Microdosing assessment to evaluate pharmacokinetics and drug metabolism using liquid chromatography-tandem mass spectrometry technology. Chapter 10 in "Topics on Drug Metabolism", edited by J. Paxton (Intech) (2012).

[23] Agilent Technologies. QTOF manual (Components) (accessed on 13th May 2014).

[24] Shimadzu. Introduction to LC-MS (Part 6). (www.shimadzu.com/an/lcms/support/intro/lib/lctalk/61/61intro.html) (accessed on 13th May 2014).

[25] M.A. Fernández-Peralbo, C. Ferreiro-Vera, F. Priego-Capote, M.D. Luque de Castro, Stable isotopic internal standard correction for quantitative analysis of hydroxyeicosatetraenoic acids (HETEs) in serum by on-line SPE–LC–MS/MS in selected reaction monitoring mode. *Talanta* 126 (2014) 170–176.

[26] M. Calderón-Santiago, J.M. Mata-Granados, F. Priego-Capote, J.M. Quesada-Gómez, M.D. Luque de Castro, Analytical platform for verification and quantitation of target peptides in human serum: application to cathelicidin. *Anal. Biochem.* 415 (2011) 39–45.

[27] M.P. Delgado-Torre, C. Ferreiro-Vera, F. Priego-Capote, P.M. Pérez-Juan, M.D. Luque de Castro, Comparison of accelerated methods for the extraction of phenolic compounds from different vino-shoot cultivars. *J. Agric. Food Chem.* 60 (2012) 3051–3060.

[28] M. Katajamaa, M. Oresic, Data processing for mass spectrometry-based metabolomics. *J. Chrom. A* 1158 (2007) 318–328.

[29] R.A. van der 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* 7:142 (2006).

[30] A. Mena-Bravo, M.D. Luque de Castro, Sweat: a sample with limited present applications and promising future in metabolomics. *J. Pharm. Biomed. Anal.* 90 (2014) 139–147.

[31] J. Xia, D.I. Broadhurst, M. Wilson, D.S. Wishart, Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* 9 (2013) 280–299.

[32] D.L. Massart, B.G.M. Vandeginste, L.M.C Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke (Eds.), Handbook of chemometrics and qualimetrics, Part A, Elsevier, Oxford, UK 1997.

[33] L. Eriksson, H. Anti, J. Gottfries, E. Holmes, E. Johansson, F. Lindgren, I. Long, T. Lundstedt, J. Trygg, S. Wold, Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal. Bioanal. Chem.* 380 (2004) 419–429.

[34] H. Hotelling, Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24 (1933) 417–441.

[35] K. Pearson, On lines and planes of closest fit to systems of points in space. *Philos. Mag*. 2 (1901) 559–572.

[36] J. Boccard, J.L. Veuthey, S. Rudaz, Knowledge discovery in metabolomics: an overview of MS data handling. *J. Sep. Sci.* 33 (2010) 290–304.

[37] H. Idborg-Björkman, P.O. Edlund, O.M. Kvalheim, I. Schuppe-Koistinen, S.P. Jacobsson, Screening of biomarkers in rat urine using LC/electrospray ionization-MS and two-way data analysis. *Anal. Chem*. 75 (2003) 4784-4792.

[38] H.J. Major, R. Williams, A.J. Wilson, I.D. Wilson, A metabonomic analysis of plasma from Zucker rat strains using gas chromatography/mass spectrometry and pattern recognition. *Rapid Commun. Mass Spectrom*. 20 (2006) 3295–3302.

[39] K.M. Pierce, J.L. Hope, J.C. Hoggard, R.E. Synovec, A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC–TOFMS) separations of metabolites in plant samples. *Talanta* 70 (2006) 797–804.

[40] D. Bylund, J. Samskog, K.E. Markides, S.P. Jacobsson, Classification of lactate dehydrogenase of different origin by liquid chromatography-mass spectrometry and multivariate analysis. *J. Am. Soc. Mass Spectrom.* 14 (2003) 236–240.

[41] P.S. Belton, I.J. Colquhoun, E.K. Kemsley, I. Delgadillo, P. Roma, M.J. Dennis, M. Sharman, E. Holmes, J.K. Nicholson, M. Spraul, Application of chemometrics to the 1H NMR spectra of apple juices: discrimination between apple varieties. *Food Chem.* 61 (1998) 207–213.

[42] J.J. Jansen, H.C.J. Hoefsloot, H.F.M. Boelens, J. van der Greef, A.K. Smilde, Analysis of longitudinal metabolomics data. *Bioinformatics* 20 (2004) 2438–2446.

[43] E.C.Y. Chan, S.L. Yap, A.J. Lau, P.C. Leow, D.F. Toh, H.L. Koh, UPLC/TOFMS based metabolomics of raw and steamed *Panax notoginseng*. *Rapid Commun. Mass Spectrom.* 21 (2007) 519–528.

[44] N. Kettaneh-Wold, Analysis of mixture data with partial least squares. *Chemomet. Intell. Lab. Syst.* 14 (1992) 57–69.

[45] S. Wold, A. Ruhe, H. Wold, W.J. Dunn, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* 5 (1984) 735–743.

[46] M. Daszykowski, B. Walczak, D.L. Massart, Projection methods in chemistry. *Chemomet. Intell. Lab. Syst.* 65 (2003) 97–112.

[47] P. Jonsson, S.J. Bruce, T. Moritz, J. Trygg, M. Sjöström, R. Plumb, J. Granger, E. Maibaum, J.K. Nicholson, E. Holmes, H. Antti, Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst* 130 (2005) 701–707.

[48] M.C. Wiener, J.R. Sachs, E. Deyanova, N.A. Yates, Differential mass spectrometry: a label-free LC–MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.* 76 (2004) 6085–6096.

[49] J.P. Shaffer, Multiple hypothesis testing. *Ann. Rev. Psychol.* 46 (1995) 561–584.

[50] O. Vorst, C.H.R. De Vos, A. Lommen, R.V. Staps, R.G.F. Visser, R.J. Bino, R.D. Hall, A non-directed approach to the differential analysis of multiple LC-MS-derived metabolic profiles. *Metabolomics* 1 (2005) 169–180.

[51] N.J. Serkova, M. Jackman, J.L. Brown, T. Liu, R. Hirose, J.P. Roberts, J.J. Maher, C.U. Niemann, Metabolic profiling of livers and blood from obese Zucker rats. *J. Hepatol.* 44 (2006) 956–962.

[52] N.A. Obuchowski, L.M. Lieber, F.H. Wians, ROC curves in clinical chemistry: uses, minuses, and possible solutions. *Clin. Chem.* 50 (2004) 118–1125.

[53] K. Soreide, Receiver-operatinf characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *J. Clin. Pathol.* 62 (2009) 1–5.

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

[55] T. Maier, M. Leibundgut, N. Ban, The crystal structure of a mammalian fatty acid synthase. *Science* 321 (2008) 1315–1322.

[56] A. D'Alessandro, B. Giardina, F. Gevi, A. M. Timperio, L. Zolla, Clinical Metabolomics: the next stage of clinical biochemistry. *Blood Transfus.* 10 (2012) 19–24.

[57] N. Friedrich, Metabolomics in diabetes research. *J. Endocrinol.* 215 (2012) 29–42.

[58] A. Zhang, H. Sun, X. Wang, Recent advances in metabolomics in neurological disease, and future perspectives. *Anal. Bioanal. Chem.* 405 (2013) 8143–8150.

[59] R. Priori, R. Scrivo, J. Brandt, M. Valerio, L. Casadei, G. Valesini, C. Manetti, Metabolomics in rheumatic diseases: the potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmunity Rev.* 12 (2013) 1022–1030.

[60] V. Fanos, C. Fanni, G. Ottonello, A. Noto, A. Dessì, M. Mussap, Metabolomics in adult and pediatric nephrology. *Molecules* 18 (2013) 4844–4857.

[61] Y.Y. Zhao, Metabolomics in chronic kidney disease. *Clin. Chim. Acta* 422 (2013) 59–69.

[62] D. Beyoglu, J. R. Idle, The metabolomic window into hepatobiliary disease. *J. Hepatol.* 59 (2013) 842–858.

[63] L.V. Hooper, T. Midtvedt, J.I. Gordon, How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Ann. Rev. Nutr.* 22 (2002) 283–307.

[64] F.P. Martin, N. Sprenger, I. Montoliu, S. Rezzi, S. Kochhar, J.K. Nicholson, Dietary modulation of gut functional ecology studied by fecal metabonomics. *J. Proteome Res.* 9 (2010) 5284–5295.

[65] F.P. Martin, I. Montoliu, S. Kochhar, S. Rezzi, Chemometric strategy for modeling metabolic biological space along the gastrointestinal tract and assessing microbial influences. *Anal. Chem.* 82 (2010) 9803–9811.

[66] Y. Wang, E. Holmes, E.M. Comelli, G. Fotopoulos, G. Dorta, H. Tang, M.J. Rantalainen, J.C. Lindon, I.E. Corthésy-Theulaz, L.B. Fay, S. Kochhar, J.K. Nicholson, Topographical variation in metabolic signatures of human gastrointestinal biopsies revealed by high-resolution magic-angle spinning HNMR spectroscopy. *J. Proteome Res.* 6 (2007) 3944–3951.

[67] Y. Wang, H. Tang, E. Holmes, J.C. Lindon, M.E. Turini, N. Sprenger, G. Bergonzelli, L.B. Fay, S. Kochhar, J.K. Nicholson, Biochemical characterization of rat intestine development using high-resolution magic-angle-spinning 1H NMR spectroscopy and multivariate data analysis. *J. Proteome Res.* 4 (2005) 1324–1329.

[68] F.P. Martin, Y. Wang, I.K. Yap, N. Srenger, J.C. Lindon, S. Rezzi, S. Kochhar, E. Holmes, J.K. Nicholson, Topographical variation in murine intestinal metabolic profiles in relation to microbiome speciation and functional ecological activity. *J. Proteome Res.* 8 (2009) 3464–3474.

[69] F.P. Martin, Y. Wang, N. Sprenger, E. Holmes, J.C. Lindon, S. Kochhar, J.K. Nicholson, Effects of probiotic *Lacto-bacillus paracasei* treatment

on the host gut tissue metabolic profiles probed via magic-angle-spinning NMR spectroscopy. *J. Proteome Res.* 6 (2007) 1471–1481.

[70] M.D. Collins, G.R. Gibson, Probiotics, prebiotics, and synbiotics: approaches for modulating the microbial ecology of the gut. *Am. J. Clin. Nutr.* 69 (1999) 1052–1057.

[71] J.A. Madden, J.O. Hunter, A review of the role of the gut microflora in irritable bowel syndrome and the effects of probiotics. *Br. J. Nutr.* 88 (2002) 67–72.

[72] S.J. Bickston, L.W. Comerford, F. Cominelli, Future therapies for inflammatory bowel disease. *Curr. Gastroenterol. Rep.* 5 (2003) 518–523.

[73] J.F.P. Martin, S. Collino, S. Rezzi, S. Kochhar, Metabolomic applications to decipher gut microbial metabolic influence in health and disease. *Front. Physiol.* 3 (2012) eCollection: doi 10.3389/fphys.2012.00113.

[74] J.T. Bjerrum, O.H. Nielsen, F. Hao, H. Tang, J.K. Nicholson, Y. Wang, J. Olsen, Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology. *J. Proteome Res.* 9 (2010) 954–962.

[75] J. van Duynhoven, E.E. Vaughan, F. van Dorsten, V. Gómez-Roldán, R. de Vos, J. Vervoort, J.J. van der Hooft, L. Roger, R. Draijer, D.M. Jacobs, Interactions of black tea polyphenols with human gut microbiota: implications for gut and cardiovascular health. *Am. J. Clin. Nutr.* 98 (2013) 1631–1641.

[76] A. Santoro, E. Pini, M. Scurti, G. Palmas, A. Berendsen, A. Brzozowska, B. Pietruszka, A. Szczecinska, N. Cano, N. Meunier, C.P. de Groot, E. Feskens, S. Fairweather-Tait, S. Salvioli, M. Capri, P. Brigidi, C. Franceschi, Combating inflammaging through a Mediterranean whole diet approach: The NU-AGE project's conceptual framework and design. *Mech. Ageing Dev.* (2013) doi: 10.1016/j.mad.2013.12.001.

[77] E. Trushina, M.M. Mielke, Recent advances in the application of metabolomics to Alzheimer's disease. *Biochim. Biophys. Aca* (2013) doi: 10.1016/j.bbadis.2013.06.014.

[78] C.B. Newgard, J. An, J.R. Bain, M.J. Muehlbauer, R.D. Stevens, L.F. Lien, A.M. Hagg, S.H. Shah, M. Arlotto, C.A. Slentz, J. Rochon, D. Gallup, O. Ilkayeva, B.R. Wenner, W.E. Yancy, H. Eisenson, G. Musante, R. Surwit, D.S. Millington, M.D. Butler, L.P. Svetkey, A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab*. 9 (2009) 311–326.

[79] K. Suhre, C. Meisinger, A. Döring, E. Altmaier, P. Belcredi, C. Gieger, D. Chang, M.V. Milburn, W.E. Gall, K.M. Weinberger, H.W. Mewes, M.H. de Angelis, H.E. Wichmann, F. Kronenberg, J. Adamski, T. Illing, Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 5 (2010) e13953.

[80] X. Zhao, J. Fritsche, J. Wang, J. Chen, K. Rittig, P. Schmitt-Kopplin, A. Fritsche, H.U. Häring, E.D. Schleicher, G. Xu, R. Lehmann, Metabonomic fingerprints of fasting plasma and spot urine reveal human pre-diabetic metabolic traits. *Metabolomics* 6 (2010) 362–374.

[81] F.J. Zieve, M.F. Kalin, S.L. Schwartz, M.R. Jones, W.L. Bailey, Results of the glucose-lowering effect of WelChol study (GLOWS): a randomized, double-blind, placebo-controlled pilot study evaluating the effect of colesevelam hydrochloride on glycemic control in subjects with type 2 diabetes. *Clin. Ther*. 29 (2007) 74–83.

[82] K. Kondo, T. Kadowaki, Colestilan monotherapy significantly improves glycaemic control and LDL cholesterol levels in patients with type 2 diabetes: a randomized double-blind placebo-controlled study. *Diabetes Obes. Metab*. 12 (2010) 246–251.

[83] P. Spegel, A.P.H. Danielsson, K. Bacos, C.L.F. Nagorny, T. Moritz, H. Mulder, K. Filipsson, Metabolomic analysis of a human oral glucose

tolerance test reveals fatty acids as reliable indicators of regulated metabolism. *Metabolomics* 6 (2010) 56–66.

[84] M. Basanta, B. Ibrahim, R. Dockry, D. Douce, M. Morris, D. Singh, A. Woodcock, S.J. Fowler, Exhaled volatile organic compounds for phenotyping chronic obstructive pulmonary disease: a cross-sectional study. *Respir. Res.* 13 (2012) doi: 10.1186/1465-9921-13-72.

[85] S. Schleidgen, C. Klingler, T. Bertram, W.H. Rogowski, G. Marckmann, What is personalized medicine: sharpening a vague term based on a systematic literature review. *BMC Med. Ethics* 14 (2013) doi: 10.1186/1472-6939-14-55.

[86] Y. Zhang, Y. Wu, T. Chen, L. Yao, J. Liu, X. Pan, Y. Hu, A. Zhao, G. Xie, W. Jia, Assessing the metabolic effects of aromatherapy in human volunteers. *Evid. Based Complement. Alternat. Med.* 2013 (2013), article ID 356381.

[87] M. Oresic, A. Vidal-Puig, V. Hänninen, Metabolomic approaches to phenotype characterization and applications to complex diseases. *Expert Rev. Mol. Diagn.* 6 (2006) 575–585.

[88] Biomarkers Definitions Working Group, Biomarkers and surrogate endpoints; preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69 (2001) 89–95.

[89] U. Pinder, Epihanie Medicorum. Speculum videndi urinas hominum. Clavis aperiendi poras pulsuum. Berillus discernendi causas & differentias febrium. Sodalitas Celtica: Nuremberg (1506).

[90] R.S. Vasan, Biomarkers of cardiovascular disease: molecular basis and practical considerations. *Circulation* 113 (2006) 2335–2362.

[91] L.K. Newby, R.H. Christenson, E.M. Ohman, P.W. Armstrong, T.D. Thompson, K.L. Lee, C.W. Hamm, H.A. Katus, C. Cianciolo, C.B. Granger, E.J. Topol, R.M. Califf, Value of serial troponin T measures for

early and late risk stratification in patients with acute coronary syndrome. *Circulation* 98 (1998) 1853–1859.

[92] Nobeli I, Thornton JM, A bioinformatician's view of the metabolome. *Bioessays*. 28 (2006) 534–545.

[93] R. Rousseau, B. Govaerts, M. Verleysen, B. Boulanger, Comparison of some chemometric tools for metabonomics biomarker identification. *Chemometr. Intell. Lab*. 91 (2008) 54–66.

[94] P.Y. Müller, F. Dieterle, Tissue-specific, non-invasive toxicity biomarkers: translation from preclinical safety assessment to clinical safety monitoring. *Expert Opin. Drug Metab. Toxicol*. 5 (2009) 1023–1038.

[95] J.A. Wagner, S.A. Williams, C.J. Webster, Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clin. Pharmacol. Ther.* 81 (2007) 104–107.

[96] G.J. Burckart, S. Amur, F.M. Goodsaid, L.J. Lesko, F.W. Frueh, S.M. Huang, M.W. Cavaille-Coll, Qualification of biomarkers for drug development in organ transplantation. *Am. J. Transplant*. 8 (2008) 267–270.

[97] J.W. Lee, R.S. Weiner, J.M. Sailstad, R.R. Bowsher, D.W. Knuth, P.J. O'Brien, J.L. Fourcroy, R. Dixit, L. Pandite, R.G. Pietrusko, H.D. Soares, V. Quarmby, O.L. Vesterqvist, D.M. Potter, J.L. Witliff, H.A. Fritche, T. O'Leary, L. Perlee, S. Kadam, J.A. Wagner, Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. *Pharm. Res*. 22 (2005) 499–511.

[98] Code Federal Regulations, Title 21, Food and drugs chapter i- -Food and Drug Administration, Department of Health and Human Services. Subchapter h, Medical Devices. Part 807, Establishment registration and device listing for manufacturers and initial importers of devices. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=807 (accessed on 4th November 2011).

[99] [accessed 11-4-2011] Code Federal Regulations, Title 21, Food and drugs chapter i- -Food and Drug Administration, Department of Health and Human Services. Subchapter h, Medical Devices. Part 814, Premarket approval of medical devices. http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=814 (accessed on 4th November 2011).

[100] Ministerial ordinance on standards for manufacturing control and quality control for medical devices and in-vitro diagnostic reagents. Pharmaceuticals and Medical Devices Agency; Tokyo: http://meddevmirq.com/Documents/MiRQ%20Overview%20of%20Japanese%20Regulatory%20System_PAL%20Summary%202009_07_18%20Sec%202.pdf (accessed on 4th November 2011).

[101] F. Dati, The new European directive on in vitro diagnostics. *Clin. Chem. Lab. Med.* 41 (2003) 1289–1298.

[102] M.E. Dumas, R.H. Barton, A. Toye, O. Cloarec, C. Blancher, A. Rothwell, J. Fearnside, R. Tatoud, V. Blanc, J.C. Lindon, S.C. Mitchell, E. Holmes, M.I. McCarthy, J. Scott, D. Gauguier, J.K. Nicholson, Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U.S.A.* 103 (2006) 12511–12516.

[103] J.H. Faber, D. Malmodin, H. Toft, A.D. Maher, D. Crockford, E. Holmes, J.K. Nicholson, M.E. Dumas, D. Baunsgaard, Metabonomics in diabetes research. *J. Diabetes Sci. Technol.* 1 (2007) 549–557.

[104] T.J. Wang, M.G. Larson, R.S. Vasan, S. Cheng, E.P. Rhee, E. McCabe, G.D. Lewis, C.S. Fox, P.F. Jacques, C. Fernandez, C.J. O'Donnell, S.A. Carr, V.K. Mootha, J.C. Florez, A. Souza, O. Mellander, C.B. Clish, R.E. Gerszten, Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17 (2011) 448–453.

[105] E. Goodman, L.M. Dolan, J.A. Morrison, S.R. Daniels, Factor analysis of clustered cardiovascular risk in adolescence: obesity is the predominant correlate risk among youth. *Circulation* 111 (2005) 1970–1977.

[106] S. Musaad, E.N. Haynes, Biomarkers of obesity and subsequent cardiovascular events. *Epidemiol. Rev.* 29 (2007) 98–114.

[107] L.Z. Yi, D.L. Yuan, Z.H. Che, Y.Z. Liang, Z.G. Zhou, H.Y. Gao, Y.M. Wang, Plasma fatty acid metabolic profile coupled with uncorrelated linear discriminant analysis to diagnose and biomarker screening of type 2 diabetes and type 2 diabetic coronary heart diseases. *Metabolomics* 4 (2008) 30–38.

[108] T. Senn, S.L. Hazen, W.H.W. Tang, Translating metabolomics to cardiovascular biomarkers. *Prog. Cardiovasc. Dis.* 55 (2012) 70–76.

[109] G.D. Lewis, L. Farrell, M.J. Wood, M. Martinovic, Z. Arany, G.C. Rowe, A. Souza, S. Cheng, E.L. McCabe, E. Yang, X. Shi, R. Deo, F.P. Roth, A. Asnani, E.P. Rhee, D.M. Systrom, M.J. Semigran, R.S. Vasan, S.A. Carr, T.J. Wang, M.S. Sabatine, C.B. Clish, R.E. Gerszten, Metabolic signatures of exercise in human plasma. *Sci. Transl. Med.* 2 (2010) 33–37.

[110] G.D. Lewis, R. Wei, E. Liu, E. Yang, X. Shi, M. Martinovic, L. Farrell, A. Asnani, M. Cyrille, A. Ramanathan, O. Shaham, G. Berriz, P.A. Lowry, I.F. Palacios, M. Tasan, F.P. Roth, J. Min, C. Baumgartner, H. Keshishian, T. Addona, V.K. Mootha, A. Rosenzweig, S.A. Carr, M.A. Fifer, M.S. Sabatine, R.E. Gerszten, Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury. *J. Clin. Invest.* 118 (2008) 3503–3512.

[111] M.S. Sabatine, E. Liu, D.A. Morrow, E. Heller, R. McCarroll, R. Wiegand, G.F. Berriz, F.P. Roth, R.E. Gerszten, Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation* 112 (2005) 3868–3875.

[112] S.J. Nicholls, Z. Wang, R. Koeth, B. Levison, B. DelFraino, V. Dzavik, O.W. Griffith, D. Hathaway, J.A. Panza, S.E. Nissen, J.S. Hochman, S.L.

Hazen, Metabolic profiling of arginine and nitric oxide pathways predicts hemodynamic abnormalities and mortality in patients with cardiogenic shock after acute myocardial infarction. *Circulation* 116 (2007) 2315–2324.

[113] W.H. Tang, Z. Wang, L. Cho, D.M. Brennan, S.L. Hazen, Diminished global arginine bioavailability and increased arginine catabolism as metabolic profile of increased cardiovascular risk. *J. Am. Coll. Cardiol.* 53 (2009) 2061–2067.

[114] Z. Wang, E. Klipfell, B.J. Bennett, R. Koeth, B.S. Levison, B. Dugar, A.E. Feldstein, E.B. Britt, X. Fu, Y.M. Chung, Y. Wu, P. Schauer, J.D. Smith, H. Allayee, W.H. Tang, J.A. DiDonato, A.J. Lusis, S.L. Hazen, Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472 (2011) 57–63.

[115] Z. Wang, W.H. Tang, L. Cho, D.M. Brennan, S.L. Hazen, Targeted metabolomic evaluation of arginine methylation and cardiovascular risks: potential mechanisms beyond nitric oxide synthase inhibition. *Arterioscler. Thromb. Vasc. Biol.* 29 (2009) 1383–1391.

[116] I. Andreadou, M. Papaefthimiou, A. Zira, M. Constatinou, F. Sigala, A.L. Skaltsounis, A. Tsantili-Kakoulidou, E.K. Iliodromitis, D.T. Kremastinos, E. Mikros, Metabonomic identification of novel biomarkers in doxorubicin cardiotoxicity and protective effect of the natural antioxidant oleuropein. *NMR Biomed.* 22 (2009) 585–592.

[117] Z. Shao, Z. Wang, K. Shrestha, A. Thakur, A.G. Borowski, W. Sweet, J.D. Thomas, C.S. Moravec, S.L. Hazen, W.H. Tang, Pulmonary hypertension associated with advanced systolic heart failure: dysregulated arginine metabolism and importance of compensatory dimethylarginine dimethylaminohydrolase-1. *J. Am. Coll. Cardiol.* 59 (2012) 1150–1158.

[118] J.S. Hochman, Cardiogenic shock complicating acute myocardial infarction: expanding the paradigm. *Circulation* 107 (2003) 2998–3002.

[119] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics. *CA Cancer J. Clin.* 61 (2011) 69–90.

[120] A. Jemal, M.M. Center, C. de Santis, E.M. Ward, Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol. Biomarkers Prev.* 19 (2010) 1893–1907.

[121] D.E. Brenner, D.P. Normolle, Biomarkers for cancer risk, early detection, and prognosis: the validation conundrum. *Cancer Epidemiol. Biomarkers Prev.* 16 (2007) 1918–1920.

[122] M. Castro, L. Grau, P. Puerta, L. Gimenez, J. Venditti, S. Quadrelli, M. Sánchez-Carbayo, Multiplexed methylation profiles of tumor suppressor genes and clinical outcome in lung cancer. *J. Transl. Med.* 8 (2010) 86–90.

[123] K.L. Richards, B. Zhang, M. Sun, W. Dong, J. Churchill, L.L. Bachinski, C.D. Wilson, K.A. Baggerly, G. Yin, D.N. Hayes, I.I. Wistuba, R. Krahe, Methylation of the candidate biomarker TCF21 is very frequent across a spectrum of early-stage nonsmall cell lung cancers. *Cancer* 117 (2011) 606–617.

[124] B. Schmidt, V. Liebenberg, D. Dietrich, T. Schelgel, C. Kneip, A. Seegebarh, N. Flemming, S. Seemann, J. Distler, J. Lewin, R. Tetzner, S. Weickmann, U. Wille, T. Liloglou, O. Raji, M. Walshaw, M. Fleischhacker, C. Witt, J.K. Field, SHOX2 DNA methylation is a biomarker for the diagnosis of lung cancer based on bronchial aspirates. *BMC Cancer* 10 (2010) 600–606.

[125] P.P. Massion, Y. Zou, H. Uner, P. Kiatsimkul, H.J. Wolf, A.E. Baron, T. Byers, S. Jonsson, S. Lam, F.R. Hirsch, Y.E. Miller, W.A. Franklin, M. Varella-Garcia, Recurrent genomic gains in preinvasive lesions as a biomarker of risk for lung cancer. *PLoS One* 4 (2009) e5611.

[126] T. Blomquist, E.L. Crawford, D. Mullins, Y. Yoon, D.A. Hernández, S. Khuder, Pattern of antioxidant and DNA repair gene expression in

normal airway epithelium associated with lung cancer diagnosis. *Cancer Res.* 69 (2009) 8629–8635.

[127] M.D. Wilkerson, X. Yin, K.A. Hoadley, Y. Liu, M.C. Hayward, C.R. Cabanski, Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin. Cancer Res.* 16 (2010) 4864–4875.

[128] N. Yanaihara, N. Caplen, E. Bowman, M. Seike, L. Kumamoto, M. Yi, Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 9 (2006) 189–198.

[129] K. Yanagisawa, S. Tomida, Y. Shimada, Y. Yatabe, T. Mitsudomi, T. Takahashi, A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. *J. Natl. Cancer Inst.* 99 (2007) 858–867.

[130] S.M.J. Rahman, A.L. González, M. Li, E.H. Seeley, L.J. Zimmerman, X.J. Zhang, Lung cancer diagnosis from proteomic analysis of preinvasive lesions. *Cancer Res.* 71 (2011) 3009–3017.

[131] K. Kami, T. Fujimori, H. Sato, M. Sato, H. Yamamoto, Y. Ohashi, N. Sugiyama, Y. Ishihama, H. Onozuka, A. Ochiai, H. Esumi, T. Soga, M. Tomita, Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. *Metabolomics* 9 (2013) 444–453.

[132] S. Hanash, Harnessing immunity for cancer marker discovery. *Nat. Biotechnol.* 21 (2003) 37–38.

[133] S.M. Hanash, S.J. Pitteri, V.M. Faca, Mining the plasma proteome for cancer biomarkers. *Nature* 452 (2008) 571–579.

[134] M.Y. Lim, P.S. Thomas, Biomarkers in exhaled breath condensate and serum of chronic obstructive pulmonary disease and non-small-cell lung cancer. *Intern. J. Chronic Dis.* 2013 (2013) Article ID 578613.

[135] P.J. Mazzone, Analysis of volatile organic compounds in the exhaled breath for the diagnosis of lung cancer. *J. Thorac. Oncol.* 3 (2008) 774–780.

[136] P. Fuchs, C. Loeseken, J.K. Schubert, W. Miekisch, Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer* 126 (2010) 2663–2670.

[137] A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, K. Schwarz, M. Ligor, Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC Cancer* 9 (2009) 348–356.

[138] M. Ligor, T. Ligor, A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry. *Clin. Chem. Lab. Med.* 47 (2009) 550–560.

[139] G. Peng, M. Hakim, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, U. Tisch, H. Haick, Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br. J. Cancer* 103 (2010) 542–551.

[140] V.S. Kurova, A.S. Kononikhin, D.A. Sakharov, I.A. Popov, I.M. Larina, A.G. Tonevitskii, S.D. Varfolomeev, E.N. Nikolaev, Exogenic proteins in the human exhaled breath condensate. *Bioorg. Khim.* 37 (2011) 55–60.

[141] C. Gessner, B. Rechner, S. Hammerschidt, H. Kuhn, G. Hoheisel, U. Sack, P. Ruschpler, H. Wirtz, Angiogenic markers in breath condensate identify non-small cell lung cancer. *Lung Cancer* 68 (2010) 177–184.

[142] I. Horvath, Z. Lazar, N. Gyulai, M. Kollai, G. Losonczy, Exhaled biomarkers in lung cancer. *Eur. Respir. J.* 34 (2009) 261–275.

[143] D.A. Pirman, E. Efuet, X.P. Ding, Y. Pan, L. Tan, S.M. Fischer, R.N. DuBois, P. Yang, Changes in cancer cell metabolism revealed by direct sample analysis with MALDI mass spectrometry. *PLoS One* 8 (2013) e61379.

[144] M. Zieba, D. Nowak, M. Suwalski, G. Piasecka, I. Grzelewska-Rzymowska, K. Tymiska, Enhanced lipid peroxidation in cancer tissue homogenates in non-small cell lung cancer. *Monaldi Arch. Chest. Dis.* 56 (2001) 110–114.

[145] B.S. Khyshiktyev, N.A. Khyshiktueva, V.N. Ivanov, S.D. Darenskaia, S.V. Novikov, Diagnostic value of investigating exhaled air condensate in lung cancer. *Vopr. Onkol.* 40 (1994) 161–164.

[146] H.P. Chan, V. Tran, C. Lewis, P. Thomas, Markers of oxidative stress in exhaled breath of subjects with lung cancer. *Respirology* 13 (2008) A58.

[147] S. Hori, S. Nishiumi, K. Kobayashi, M. Shinohara, Y. Hatakeyama, Y. Kotani, N. Hatano, Y. Maniwa, W. Nishio, T. Bamba, E. Fukusaki, T. Azuma, T. Takenawa, Y. Nishimura, M. Yoshida, A metabolomic approach to lung cancer. *Lung Cancer* 74 (2011) 284–292.

[148] M. Calderón-Santiago, F. Priego-Capote, B. Jurado-Gámez, M.D. Luque de Castro, Optimization study for metabolomics analysis of human sweat by liquid chromatography–tandem mass spectrometry in high resolution mode. *J. Chromatogr. A* 1333 (2014) 70–78.

[149] Not published: M. Calderón-Santiago, F. Priego-Capote, N. Turck, X. Robin, B. Jurado-Gámez, J.C. Sanchez, M.D. Luque de Castro, Human sweat metabolomics for lung cancer prediction, personal communication.

[150] B. Jurado-Gámez, A. Salvatierra-Velázquez, M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro, Método de clasificación, diagnóstico y seguimiento de individuos con riesgo de padecer cáncer de pulmón mediante el análisis del sudor. Patente P201331228.

[151] M. Mayr, B. Madhu, Q. Xu, Proteomics and metabolomics combined in cardiovascular research. *TCM* 17 (2007) 43–48.

HERRAMIENTAS Y EQUIPOS ANALÍTICOS

Analytical tools and equipment

En este apartado de la Memoria se describen sucintamente los diferentes instrumentos y aparatos usados durante el desarrollo experimental de la Tesis. En los capítulos se incluye una explicación más detallada de los que se han utilizado en cada uno de ellos.

1. Sistemas automáticos y/o continuos de preparación de muestra

En el Capítulo 2 de la Memoria se ha tratado de enfatizar la aplicabilidad de los sistemas continuos al tratamiento de muestra en el análisis no orientado (untargeted analysis) para aumentar el número de metabolitos detectados. Por otro lado, en el bloque 3 de la Memoria se recoge el uso de un sistema automático para la preparación de muestra en análisis orientado (targeted analysis). Estos sistemas automáticos permiten llevar a cabo de manera reproducible, parcial o totalmente automatizada —y a veces con drástica reducción del volumen de muestra y reactivos— esta etapa crucial del proceso analítico, que es una de las principales fuentes de error de los métodos de análisis cuantitativo.

Se emplearon dos sistemas comerciales similares para llevar a cabo la SPE de forma automatizada: el Prospekt-2, que se usó para el pretratamiento de muestra en el análisis de catelicidina, como se recoge en el Capítulo 11; y el Symbiosis, que se usó para el pretratamiento de muestra en el análisis de aminoácidos (Capítulo 9) y para maximizar el número de metabolitos detectados en análisis no orientado (Capítulo 2). Ambos sistemas, además de trabajar en modo dinámico, permiten la elución directa con la fase móvil cromatográfica o con un pequeño volumen de disolvente que se incorpora a la fase móvil. Trabajar a alta presión posibilita la conexión en línea de este sistema con el conjunto cromatográfo/detector, consiguiéndose así la automatización completa del método analítico. El sistema Prospekt-2 está compuesto por tres módulos: Un

muestreador (MIDAS), una unidad de extracción en fase sólida (ACE) y una bomba de alta presión dispensadora de disolventes de 2 mL de capacidad (HPD). Por su parte, el sistema Symbiosis cuenta con 5 módulos: Un muestreador refrigerado (AS), de enorme interés para muestras termoinestables; una unidad de extracción en fase sólida (ACE), dos bombas de alta presión dispensadoras de disolventes (HPD1 y HPD2), de 2 mL de capacidad, y una bomba de HPLC.

2. Sistemas discontinuos de preparación de muestra

Al igual que los del apartado anterior, los sistemas discontinuos se han empleado tanto en análisis orientado como no orientado.

Dada la naturaleza no selectiva, semi- o no cuantitativa de las plataformas analíticas empleadas a lo largo de la Tesis para análisis metabólico global, se hizo especial hincapié en el desarrollo de métodos con mínima o nula preparación de muestra, evitándose, por tanto, los sistemas continuos de pretratamiento de muestra (salvo en el estudio recogido en el Capítulo 2). En general, el pretratamiento de muestra se basó en etapas de extracción, centrifugación, hidrólisis o extracción en fase sólida usando µ-SpinColumn (un sistema de SPE diseñado para pequeños volúmenes de muestra y en el que se utiliza la fuerza centrífuga para hacer pasar los disolventes a través del cartucho). Por otro lado, para la extracción y preconcentración de fosfolípidos se usaron cartuchos de SPE convencionales, empaquetados con un sorbente específico para esta familia de compuestos (HybridSPE), según se recoge en el Capítulo 8.

La determinación de los compuestos implicados en el ciclo de los ácidos tricarboxílicos mediante GC-IT/MS requirió un pretratamiento de muestra basado en la precipitación de las proteínas del suero, la extracción y la derivatización por sililación de estos metabolitos (ver Capítulo 10).

3. Detección con o sin separación cromatográfica previa

Los métodos desarrollados en la parte experimental de esta Tesis Doctoral abarcan tanto los que requieren separación cromatográfica (LC o GC) previa a la detección mediante espectrometría de masas, como los que la detección se lleva a cabo sin separación.

En los Capítulos 9 y 11, dedicados al análisis orientado de aminoácidos y catelicidina, la separación mediante LC y posterior detección por MS en tándem por triple cuadrupolo se llevó a cabo con un cromatógrafo Agilent 1200 Series LC equipado con una bomba binaria, un desgasificador, un automuestreador y un compartimento de columna termostatizados, y un espectrómetro de masas Agilent 6410 de triple cuadrupolo con una fuente de ionización por electrospray (ESI). El software Agilent MassHunter Workstation se usó para la toma de datos y el análisis cuali- y cuantitativo. El Capítulo 9 recoge el uso de una fase estacionaria de interacción hidrófila –HILIC–, especialmente diseñada para compuestos polares o iónicos que tienen poca o ninguna retención en columnas de fase reversa, tal como requiere la separación de aminoácidos.

En las plataformas dedicadas a la obtención del perfil metabolómico (metabolomics profiling) de muestras de sudor (Capítulos 1 y 7) y suero (Capítulos del 2 al 6), y en las desarrolladas para la identificación y cuantificación de fosfolípidos en suero (Capítulo 8), se utilizó un equipo HPLC Agilent 1200 Series acoplado a un detector de masas de tiempo de vuelo de alta resolución, Agilent 6540. Las columnas para la separación cromatográfica fueron en su mayoría C18 (fase reversa), salvo en el caso de las plataformas de los Capítulos 1 y 2 en los que también se ensayaron columnas con fase estacionaria HILIC. En todos los casos se usó el software MassHunter para la adquisición de espectros y el análisis cualitativo.

Para el análisis de los compuestos implicados en el ciclo de los ácidos tricarboxílicos se empleó un cromatógrafo de gases Varian CP-3800 acoplado a un detector de masas de trampa iónica (Varian Saturn 2200), que consta de un automuestreador (Varian 8400). Se utilizó el software de Varian de control de

sistema e integración de señales (Star Chromatography Workstation 6.0) para la obtención y el tratamiento de los datos.

4. Técnicas quimiométricas

De acuerdo con la importancia que ha adquirido la quimiometría en los métodos usados en metabolómica, en la Tesis Doctoral que se recoge en esta Memoria se han utilizado extensamente herramientas quimiométricas, tanto para el desarrollo y optimización de métodos analíticos como para el tratamiento de datos multivariantes. La metodología del diseño de experimentos se ha utilizado, cuando ha sido posible, en la optimización de algunas condiciones experimentales como las implicadas en la ionización de analitos previa a su detección por espectrometría de masas. La precisión de los métodos propuestos para el análisis cuantitativo (targeted analysis) de familias de compuestos se estudió como reproducibilidad dentro del laboratorio y repetibilidad mediante series de experimentos por triplicado usando análisis de varianza (ANOVA) a diferentes niveles de concentración de los analitos.

El tratamiento de datos se realizó con distintos programas informáticos según el objetivo:

a) Alineamiento de entidades moleculares en análisis no orientado (profiling):

   - Por un lado se usó la combinación de dos paquetes del lenguaje de programación R de uso libre: XCMS y CAMERA. El primero permite extraer y alinear las entidades moleculares potenciales de un análisis no orientado, y el segundo reconoce aductos e isótopos y los agrupa para eliminar falsos positivos y crear la lista definitiva de entidades.

   - Por otro lado se empleó la combinación de dos softwares de Agilent: Qualitative Workstation y MassProfiler Professional (Agilent). El primero permite extraer las entidades teniendo en cuenta aductos e

isótopos, mientras que el segundo posibilita el alineamiento de las entidades potenciales.

- Finalmente, se usó el software MassProfinder, de Agilent, que permite realizar las etapas de extracción de entidades (teniendo en cuenta aductos e isótopos) y el alineamiento en un único paso.

b) Análisis estadístico:

- El paquete estadístico "stats" del lenguaje de programación R, que permite realizar una amplia gama de análisis estadísticos tanto paramétricos como no paramétricos, así como la representación de los datos.

- Statgraphics y Unscrambler: Ambos softwares también permiten realizar distintos análisis estadísticos univariantes y multivariantes, así como el diseño y evaluación de modelos de screening y superficies de respuesta.

- MassProfiler Professional: Permite la aplicación de diferentes algoritmos de análisis estadístico especialmente adecuados para el análisis metabolómico.

c) Evaluación de la capacidad de predicción de metabolitos o paneles con potencial marcador: Esta evaluación se ha realizado a través de las curvas ROC, que representan especificidad frente a sensibilidad. Para su obtención se ha usado tanto la herramienta gratuita de acceso online ROCCET (http://www.roccet.ca/ROCCET/), como el paquete "pROC" del lenguaje de programación R.

d) Diseño de modelos de predicción y creación de paneles de marcadores: En este caso se han empleado tanto el MassProfiler Professional que permite la creación de modelos por PLS-DA, como la herramienta en línea ROCCET o el software PanelomiX, que permiten crear paneles de marcadores y evaluar su capacidad de predicción/discriminación.

5. Bases de datos

Aunque queda un gran camino por recorrer, existen varias bases de datos de metabolitos a disposición del usuario que contienen información para la identificación y caracterización de muchos de los compuestos presentes en diversas matrices biológicas. Entre ellas, las bases de datos 'Metabolites and Tandem MS Database' (METLIN), 'Human Metabolome Database' (HMDB) y 'MassBank' se han empleado para la identificación de compuestos presentes en biofluidos (suero y sudor) a partir de los datos espectrales obtenidos mediante espectrometría de masas en las partes 1, 2 y 5 de esta Memoria. Además, la información biológica disponible en la HMDB y las rutas biosintéticas de la base de datos 'Kyoto Encyclopedia of Genes and Genomes' (KEGG) se han utilizado para la interpretación de resultados obtenidos en los capítulos 4, 5 y 6.

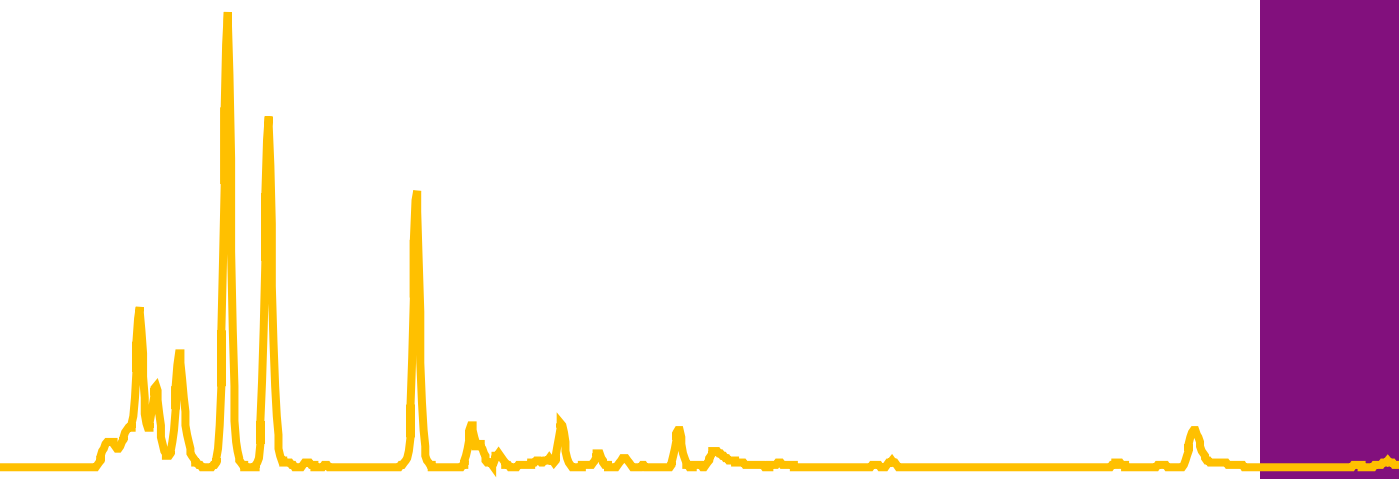
Para la identificación de lípidos existen otras bases de datos que, si bien no están constituidas por datos espectrales experimentales, contienen información útil sobre la rotura *in silico* de las distintas familias, como es el caso de LipidMaps. Esta base de datos se empleó, en las partes 3 y 5 de esta Memoria, para la identificación de fosfolípidos en base a los datos espectrales obtenidos por espectrometría de masas.

PARTE EXPERIMENTAL

Experimental part

SECTION I

Methodological development and innovation on metabolomics analysis



Section I of this PhD Book is devoted to methodological development and innovation on metabolomics with the main aim of helping to overcome some of the weak aspects of profiling analysis in clinical metabolomis. The target aspects are related both with clinical sample preparation of different biofluids involving different tools, and with coverage enhancement of metabolomics profiling.

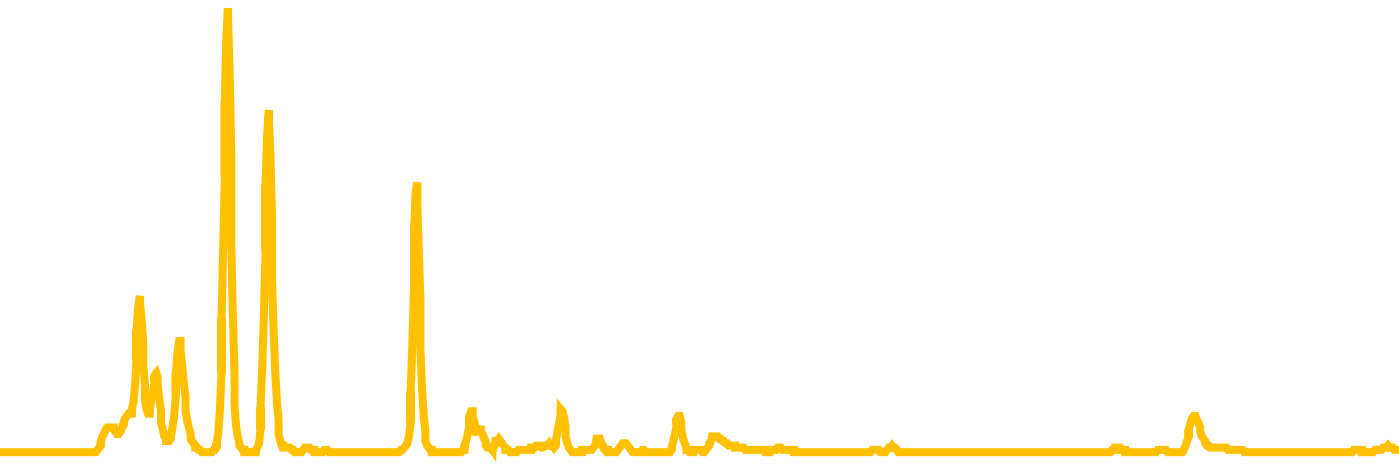
Chapter 1 is devoted to the development of a method to analyze human *sweat*, an odd clinical sample, with clear potential in metabolomics. The method, based on LC–QTOF MS/MS, provided identification of 41 metabolites after sample preparation either by only dilution or clean-up by SPE to increase sensitivity and reduce interferents. The pathways in which most of the identified metabolites are involved open a door to the use of this sample to searching for biomarkers.

The limited detection coverage of the approaches for global metabolomics profiling of *serum* has intended to be overcome in Chapter 2, in which an automated approach was configured by online coupling of SPE involving sorbents with different retention mechanisms to LC–QTOF. The SPE protocols allowed 3445 molecular entities to be detected by serial configurations of the automated SPE system.

One other limitation of metabolomics profiling is the necessity for MS/MS information to achieve unequivocal metabolite identification. The general procedure in most analytical platforms requires a first injection of the sample batch to know the potential metabolites to be identified, and a second injection to obtain the MS/MS information of the ion precursors. To avoid dual injection, Chapter 3 deals with the study of the GPF strategy advantages in profiling metabolomics analysis.

Chapter 1:

Optimization study for metabolomics analysis of human sweat by liquid chromatography–tandem mass spectrometry in high resolution mode



Journal of Chromatography A

1333 (2014) 70–78

Optimization study for metabolomics analysis of human sweat by liquid chromatography–tandem mass spectrometry in high resolution mode

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b]*, *B. Jurado-Gámez* [b,c], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Department of Respiratory Medicine, Reina Sofia University Hospital. Córdoba, Spain.*

Optimization study for metabolomics analysis of human sweat by liquid chromatography–tandem mass spectrometry in high resolution mode

M. Calderón-Santiago, F. Priego-Capote, B. Jurado-Gámez, M. D. Luque de Castro**

Abstract

Sweat has recently gained popularity as a potential tool for diagnostics and biomarker monitoring as it is a non-invasive biofluid the composition of which could be modified by certain pathologies, as is the case with cystic fibrosis, which increases chloride levels in sweat. The aim of the present study was to develop an analytical method for analysis of human sweat by liquid chromatography–mass spectrometry (LC–QTOF MS/MS) in high resolution mode. Thus, different sample preparation strategies and different chromatographic modes (HILIC and C18 reverse modes) were compared to check their effect on the profile of sweat metabolites. Forty one compounds were identified by the MS/MS information obtained with a mass tolerance window below 4 ppm. Amino acids, dicarboxylic acids and other interesting metabolites such as inosine, choline, uric acid and tyramine were identified. Among the tested protocols, direct analysis after dilution was a suited option to obtain a representative snapshot of sweat metabolome. In addition, sample clean up by C18 SpinColumn SPE cartridges improved the sensitivity of most identified compounds and reduced the number of interferents. As most of the identified metabolites are involved in key biochemical pathways, this study opens new possibilities to the use of sweat as a source of metabolite biomarkers of specific disorders.

1. Introduction

Sweat is an aqueous electrolyte solution excreted by the eccrine and apocrine sweat glands originated in the skin dermis of mammals and terminated in the secretory canals that flow into the skin surface and hair follicles. The primary function of sweating is thermoregulation to control body temperature by evaporative cooling. In addition, sweat is a defence mechanism of skin, the excretion fluid of chemosignals such as androstadienone —which acts as hormonal stimuli in females—, and waste of metabolites such as uric acid [1]. Sweat is mainly composed by water, but it contains several minor components including electrolytes, ammonia, urea, small molecules such as carboxylic acids and amino acids as well as more complex biomolecules such as proteolytic enzymes and antimicrobial peptides, among others [2]. Therefore, the varied composition of sweat supports its clinical interest to be potentially exploited for diagnostic.

The scant application of sweat in clinical analysis is explained by the lack of studies to relate sweat composition and pathological states. In fact, few clinical tests use sweat samples. One of these tests is for diagnosis of cystic fibrosis in new-borns based on determination of chloride in sweat [3–5]. However, the advances both in sweat collection devices and sensitive analytical techniques increased the interest on sweat testing of drugs over the past few years [6–10]. Sweat may offer a non-invasive alternative to present sampling for continuous monitoring of drugs exposure as sweat can be collected for a programmed period with minimal disturbance for the sampled individual. Sweat sampling can be performed with the aid of sweat wipes as liquid perspiration or over time using sweat patches [11]. The experience gained in sweat collection procedures for the analysis of xenobiotics could be used to implement this biofluid in clinical diagnostics [12,13].

Major limitations in the analysis of sweat are the low volume of sweat secreted and the variability in secretion among individuals [14,15]. Also, sampling is a complex process since the collected sweat should be representative and

unaffected by interferents. Non-invasiveness and the absence of requirements in terms of health personnel care are the main benefits from sweat as clinical sample.

Despite the scant studies on the clinical perspectives of sweat, the potential of this biofluid in omics disciplines has been pointed out. In proteomics, a current research developed by Raiszadeh et al. revealed that sweat proteome is rather different from serum proteome [16]. Therefore, human sweat could be considered an additional source of unique disease-associated biomolecules. In fact, Raiszadeh et al. found differential abundances of selected proteins between schizophrenia patients and control individuals. This preliminary test should be validated to prove the applicability of human sweat in the diagnostic of schizophrenia. In the metabolomics field, a recent study of sweat by high-resolution NMR spectroscopy has revealed that some of the components present in human sweat are involved in primary and secondary biological functions [17]. Among them, amino acids, sugars, lactate, glycerol, and compounds involved in the citric acid cycle (e.g. pyruvate, fumarate or aconitate) have been detected. NMR is especially suited to metabolomics profiling of human sweat as this is a relatively non-complex biofluid. Nevertheless, more research on sweat composition is demanded to assess the potential of this biofluid for clinical diagnostic. The aim of the present study was to develop a method for analysis of human sweat by LC–QTOF MS/MS. Different sample preparation strategies such as sample dilution, acid/alkaline hydrolysis and solid-phase extraction (SPE) were compared to check their influence on the profile of the detected metabolites. Taking into account the polar character of human sweat, two analytical chromatographic columns, C18 and HILIC, were tested to study their suitability to analyze human sweat. Identification of metabolites by LC–QTOF MS/MS in high resolution mode was carried out to obtain a snapshot of the composition of sweat metabolome.

2. Experimental

2.1. Reagents

MS-grade formic acid, ammonium formate and acetonitrile (ACN) to prepare the chromatographic mobile phases and sodium hydroxide and hydrochloric acid were purchased from Scharlab (Barcelona, Spain). Deionized water (18 mΩ • cm) from a Millipore Milli-Q water purification system was used.

2.2. Instruments and apparatus

An Agilent 1200 Series LC system (consisting of a binary pump, a vacuum degasser, an autosampler and a thermostated column compartment) coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with dual electrospray ionization (ESI) source (Santa Clara, CA, USA) was used. The chromatographic eluate was thus monitored in high resolution mode.

2.3. Cohort selected for the study

A cohort of 96 individuals with an average age of 59±11 years and a proportion of 75% male individuals participated in this study. All steps from sweat sampling to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki of 2004. 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. Sweat producer and collector: procedure

All samples were collected from 9 to 11 am after breakfast ingestion to fix the most common conditions of clinical practice. A Macroduct® Sweat Analysis System (Wescor, Utah, USA), consisting of a Webster sweat inducer and a Macroduct sweat collector (US Patent 4,542,751), was used. Pilogel® iontophoretic discs (US Patent 4,383,529) (Wescor, Utah, USA), a gel reservoir of pilocarpinium ions, were used in the iontophoretic stimulation of sweat.

The sweat inducer provided a current intensity of 1.5 mA for 5 min through two pilogel discs as electrodes located on the forearm. After removing the discs the skin where the positive disc had been located was cleaned with distilled water and the Macroduct collector covered this skin to collect sweat for 15 min. The collected sweat was transferred to a micro Eppendorf and stored at –80 ºC until use.

One individual was randomly selected to collect seven sweat samples in different days which were used to study sample variability and methodological reproducibility.

2.5. Sample treatment

A pool was prepared taking 5 μL of samples from all participants. Then, different experimental protocols were compared to select that providing maximum metabolite coverage. Firstly, 10 μL of sample was diluted with 20 μL of 0.1% v/v formic acid in water to analyze directly sweat without treatment. Then, two different strategies were tested.

(i) Sweat hydrolysis under acid or alkaline conditions to release metabolites conjugated to proteins or peptides. With this aim, 100 μL aliquots were 1:1 mixed with 0.1 M either NaOH or HCl in water and vortexed at room temperature for 30 min, then evaporated to dryness and reconstituted by 100 μL of chromatographic mobile phase A.

(ii) Sweat clean up and preconcentration by solid phase extraction using C18 and hydrophilic centrifugal Micro SpinColumn™ systems (Harvard Apparatus, MA, USA) and following the protocol recommended by the manufacturer depending on the sorbent material.

The protocol for C18 Micro SpinColumn™ was as follows: 150 μL of water for solvation, 150 μL of 50% (v/v) acetonitrile for sorbent conditioning, 150 μL of 5% (v/v) acetonitrile for sorbent equilibration, 25 μL of sample, 150 μL of 5% (v/v) acetonitrile to obtain thesample free from the undesirable retained

compounds, which were subsequently eluted by 150 μL of 50% (v/v) acetonitrile. Centrifugation for 25 s at 1000 × g was used for each step.

The protocol for hydrophilic Micro SpinColumn™ system was: 150 μL of water for solvatation, 150 μL of 50% (v/v) acetonitrile for sorbent conditioning, 150 μL of 5% (v/v) acetonitrile for sorbent equilibration, 25 μL of sample, 150 μL of acetonitrile to remove undesired non retained compounds, and 150 μL of 50% (v/v) acetonitrile to elute the retained target compounds. The centrifugation conditions were as those used for the C18 sorbent.

Both the eluate and the nonretained fraction (collected after sample application) obtained in each SPE protocol were analyzed separately after evaporation and reconstitution by 20 μL of water acidified with 0.1% formic acid.

2.6. LC–QTOF MS/MS analysis

Two chromatographic columns were tested for separation. A C18 reverse-phase analytical column (Mediterranean, 50 mm×0.46 mm i.d., 3 μm particle size) from Teknokroma (Barcelona, Spain) was used at constant temperature of 25 °C. The mobile phases were water (phase A) and ACN (phase B) both with 0.1% formic acid as ionization agent. The LC pump was programmed with a flow rate of 0.8 $mLmin^{-1}$ and the following gradient elution was developed: 3% phase B was kept constant from min 0 to 1; from 3 to 80% of phase B from min 1 to 10.5 and from 80 to 100% of phase B from min 10.5 to 11.5. A post-time of 5 min was set to stabilize the initial conditions.

A Luna hydrophilic interaction chromatography column (HILIC) (100 mm × 0.46 mm i.d., 3 μm particle size) from Phenomenex (Torrance, CA, USA) was used in the other method. In this case, the mobile phases were 5% (v/v) of ACN (phase A) and 95% (v/v) of ACN (phase B), both with 5 mM of ammonium formate (pH 6) as ionization agent. The LC pump was programmed with a flow rate of 0.6 $mLmin^{-1}$ and the following gradient elution was developed: 90% phase B was kept constant for 5 min; then, the percentage of phase B decreased to 70% since min 5 to 17 and finally decreased to 50% from min 17 to 20. A post-time of 5 min was set to stabilize the initial conditions.

In both cases the injection volume was 3 μL and the injector needle was washed for 10 times between injections with 70% methanol. Furthermore, the needle seat back was flushed for 15 s at a flow rate of 4 $mLmin^{-1}$ with 70% methanol to avoid cross contamination.

The parameters of the electrospray ionization source, operating in negative and positive ionization mode, were as follows: the capillary and fragmentor voltage were set at ±3.5 kV and 175 V, respectively; N_2 in the nebulizer was flowed at 40 psi; the flow rate and temperature of the N_2 as drying gas were 8 L min^{-1} and 350 °C, respectively. The instrument was calibrated and tuned according to the procedures recommended by the manufacturer. Data were collected in both centroid mode at a rate of 1 spectrum per second in the extended dynamic range mode (2 GHz).Accurate mass spectra in MS scan and MS/MS mode were acquired in the m/z range 60–1100. The instrument gave typical resolution 15000 FWHM (Full Width at Half Maximum) at m/z 118.0862 and 30000 FWHM at m/z 922.0098. 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 [protonated hexakis (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 of HP-921) were used.

The samples were first injected in full scan acquisition mode and then in auto MS/MS acquisition mode to obtain information on the product ion of the most relevant compounds. The maximum number of precursors selected per cycle was set at 2 with an exclusion window of 0.3 min after 2 consecutive selections of the same precursor. Three collision energies (10, 20 or 40 eV) were used to obtain the maximum information from fragmentation. At least three replicates of each sample preparation test were analyzed to improve the quality of the results. Blanks were inserted in the analysis sequences between sets of 5 samples.

2.7. Data processing and statistical analysis

MassHunter Workstation software (version B5.00 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) was used to process all data obtained by LC–QTOF in full scan MS mode. Treatment of raw data file was initiated 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 500 and 1000 counts with a single charge state for chromatograms obtained with C18 and HILIC column, respectively, as the background noise was different for the two cases. Additionally, the isotopic distribution to consider a molecular feature as valid should be defined by two or more ions (with a peak spacing tolerance of 0.0025 m/z, plus 10.0 ppm in mass accuracy). Adducts formation in the positive (+H, +Na, +NH4) and negative ionization (-H, +HCOO) modes, as well as neutral loss by dehydration were included to identify features corresponding to the same potential metabolite. Thus, ions with identical elution profiles and related m/z values (representing different adducts or isotopes of the same compound) were extracted as entities characterized by their retention time (RT), intensity in the apex of the chromatographic peaks and accurate mass. Background contribution was removed by subtraction of MFs linked to plasticizers, solvent impurities and other contaminants after analysis of a blank (0.1% formic acid) under identical operational conditions for each sample treatment procedure tested. In this way, raw data files were created in compound exchange format files (.cef files) for each sample and exported into the Mass Profiler Professional (MPP) software package (version 12.1, Agilent Technologies, Santa Clara, CA, USA) for further processing.

In the next step, the data were processed by alignment of retention times and m/z values across the data matrix using a tolerance window of 0.6 min and 10 ppm mass accuracy, respectively. Data pretreatment was based on baselining to remove background noise and normalization by logarithmic transformation to reduce relatively large differences among the respective MF abundances. The molecular features extracted from the analysis of blanks were removed from the data set formed by molecular features extracted from sweat samples. The

extraction algorithm confirmed the efficiency of this filtering step. This correction was applied to all the sample treatments tested. Stepwise reduction of the MFs number was performed based on frequency of occurrence by comparing repetitions of the same fraction or hydrolysis procedure. A filter by frequency was set at 100%, which ensured the detection of each molecular feature in all the injected replicates .

Identification of the most relevant entities was based on MS and MS/MS information and searching in the METLIN MS and MS/MS database (http://metlin.scripps.edu) and Human Metabolome Database (HMDB, 3.5 version).

3. Results and discussion

As mentioned before, sweat should be considered a potential biofluid for clinical diagnostic of great interest because its non-invasive sampling. With this aim, development of methods for analysis of sweat by LC–QTOF MS/MS are demanded to define suited protocols for characterization of this biofluid by coverage of as much as possible sweat metabolome.

3.1. Comparison of the two chromatographic modes for direct analysis of human sweat

Most metabolites present in sweat are polar because of the aqueous nature of this excreted biofluid. For this reason, the separation capability of two chromatographic modes, HILIC and C18 reverse mode, were compared. In both cases the sweat pool was 1:1 diluted with 0.1% (v/v) formic acid and directly analyzed by LC–TOF MS in positive and negative ionization modes. Molecular features were extracted from raw data files obtained by the four chromatographic methods (C18 and HILIC methods in positive and negative ionization modes). For this purpose, an extraction algorithm was used by setting the minimum peak abundance at 500 counts for C18 reverse mode analyses and 1000 counts for HILIC analyses. These cut-off values were established taking into account the

chromatographic background noise, which was higher in the HILIC chromatograms. Supplementary Fig. 1 shows the distribution of the retention times for the molecular features detected by each chromatographic method and ionization mode. As can be seen, most of the molecular features detected in the methods based on the C18 analytical column were associated to retention times in the first part of the chromatogram, while resolution in the case of HILIC allowed a more homogeneous elution of potential metabolites. Therefore, the polar character of the metabolite composition of sweat is confirmed. Fig. 1 shows the base peak chromatograms (BPCs), which included the intensity of the highest ion per scan, obtained by the HILIC (Fig. 1a) and C18 (Fig. 1b) analytical columns in both ionization modes. As can be seen, the two chromatograms reporting wider information in terms of number of peaks were obtained with the method based on the HILIC column in the negative ionization mode and with that on C18 column in the positive ionization mode. On the other hand, the chromatograms provided by HILIC in the positive ionization mode and C18 in the negative mode were not characterized by high resolution. It is worth emphasizing that BPCs were influenced by the presence of components of the gel discs. Among gel discs components to stimulate sweat secretion, the contaminant giving the highest peak in sweat samples was pilocarpine, which is mostly detected in the positive ionization mode. Other components of the gel formulation such as methylparaben and propylparaben were also detected in the analysis of sweat by both ionization modes.

Direct analysis of sweat enabled detection of 67 and 57 molecular features in positive ionization mode for the C18 and HILIC analytical methods, respectively; meanwhile, 29 and 37 molecular features, respectively, were detected in negative ionization mode for both columns. Although C18 reverse mode seems not to be appropriate for separation of sweat polar metabolites according to the chromatographic profile, this mode should not be discarded at least in positive ionization mode due to the number of molecular entities detected.

Despite the high number of potential entities detected in positive ionization mode, sample preparation could improve metabolite coverage. With this aim, different steps encompassing hydrolysis or cleaning by centrifugal SPE were tested after a quality control study to assess the performance of the analysis equipment.

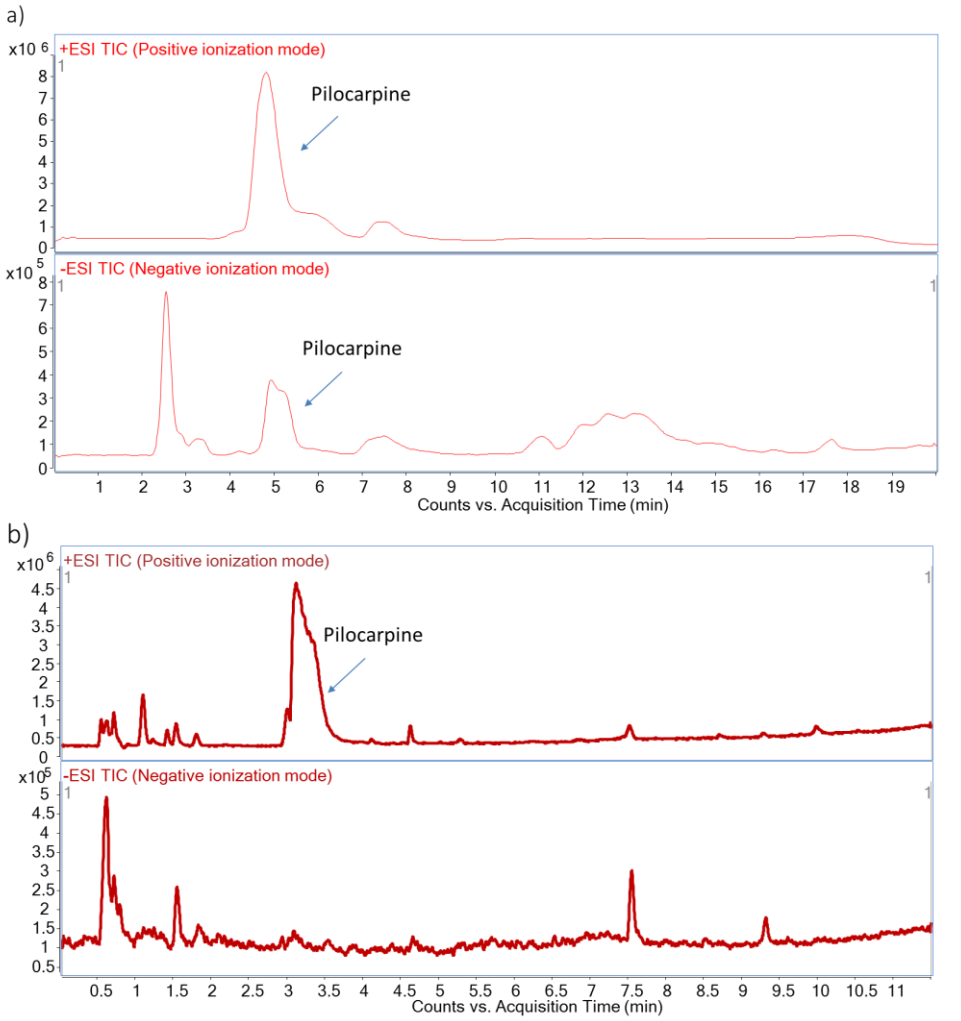


Fig. 1. Base peak chromatograms (BPCs) obtained for sweat analysis in positive and negative ionization using two chromatographic modes: (a) HILIC, (b) RP-C18.

3.2. Quality control (QC) study

A quality control (QC) test was carried out by analysis of sweat as an attempt to assess the instrumental precision of the detection step and the stability of the samples. The QC test was performed by repetitive injection (n = 14) of the human sweat pool prepared in this research, by inserting a blank each 5 injections to evaluate carry-over effects. The autosampler temperature was set at 6 ºC to minimize potential degradation and methylparaben was used as internal standard as a normalization strategy to minimize intra-individual variability. Methylparaben is always present in sweat and its concentration would depend on the sweat volume collected. Due to the number of molecular features found in the samples (67 in the positive ionization mode), a multivariate analysis based on Principal Component Analysis (PCA) was used to assess variability associated to instrumental precision and sample integrity. Supplementary Fig. 2 shows the PCA resulting from normalization in which significant variations were observed after 4 h at 6 ºC (sample number 9 of the programmed sequence in the QC test). Therefore, sequences of analysis included only 8 samples to minimize variability sources associated to sample stability. The normalization procedure was considered for programming sequences in further experiments. Under these conditions, instrumental precision estimated as RSD percentage ranged from 3 to 11%.

Additionally, the intra-individual variability was assessed by analysis of 7 samples collected from the same individual in different days, which were analyzed in duplicate within a five days interval. This study also allowed evaluating the methodological reproducibility since the variability estimated as RSD ranged from 2 to 21%. These values indicated that the collection system was reproducible enough for metabolomics research.

3.3. Comparison of acidic and basic hydrolysis

Sweat hydrolysis protocols under acidic and alkaline conditions were tested to maximize metabolite identification in sweat as some metabolites could thus be released from proteins or other conjugates. For this purpose, sweat was

incubated with NaOH or HCl aqueous solutions as described in the Experimental section. Fig. 2a and 2b shows the chromatograms from sweat without hydrolysis and after basic and acid hydrolysis in positive and negative ionization modes, respectively, using the C18 analytical column. As can be seen, the number of chromatographic peaks increased significantly by analysis of hydrolyzed samples as compared to direct analysis of sweat. Fig. 2c and 2d compares the number of features detected in all the repetitions in each analysis in negative and positive ionization modes, in which the number of features increased significantly in the negative ionization mode for sweat hydrolyzed under alkaline conditions, while in the positive mode the features increased under acidic conditions. Thus, the analysis of sweat subjected to alkaline hydrolysis allowed detection of 155 potential entities in negative ionization mode, while direct analysis of sweat led to 29 entities. On the other hand, acidic hydrolysis enabled detection of 83 entities in positive ionization mode, while direct analysis led to detection of 67 entities. One fact to be emphasized is that 79 and 76% of the molecular features detected by direct analysis of sweat were also detected in the hydrolyzed samples in negative and positive ionization modes, respectively. Despite the high number of potential metabolites detected in hydrolyzed sweat, many of them were identified as pilocarpine adducts (sodium, dimers and fragments), the main contaminant interferent detected in human sweat, which is used to stimulate sweat secretion by the Macroduct® system. In fact, 8 out of 11 molecular features exclusively detected in the analysis of sweat subjected to alkaline hydrolysis in positive ionization mode were identified as pilocarpine adducts, and 15 out of 26 entities exclusively detected in sweat subjected to acid hydrolysis in positive ionization mode were identified as pilocarpic acid adducts or fragments. However, only 7 out of 122 molecular features exclusively detected in sweat after alkaline hydrolysis were identified as pilocarpine adducts in negative ionization mode. In this case, peptide fragments and artifacts were among the numerous molecular features detected in sweat hydrolyzed under alkaline conditions. Attending to these results, sweat hydrolysis would not be recommended since the sample

integrity is altered by detection of numerous artifacts associated to metabolites and peptides.

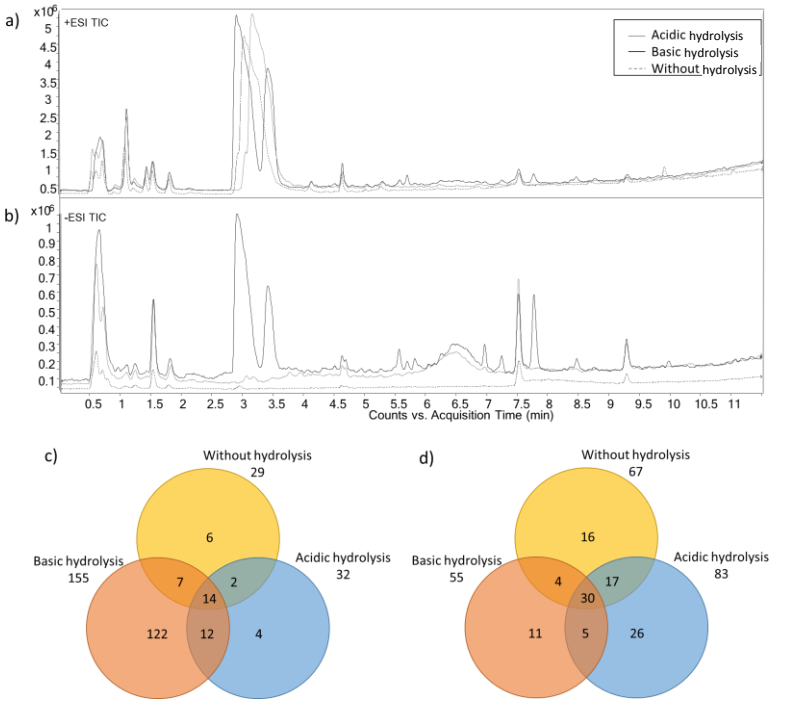


Fig. 2. BPCs obtained by analysis of human sweat without hydrolysis and after alkaline and acid hydrolysis in positive (a) and negative (b) ionization modes. Venn diagrams comparing the molecular features detected in the three analyses in negative (a) and positive (b) ionization modes.

3.4. Sample cleaning by centrifugal solid phase extraction

The Macroduct® sweat collection system is commercialized for the determination of chloride in sweat as a diagnostic tool for cystic fibrosis. It is easy to use, reproducible, and sample collection does not require an additional extraction procedure. However, this system is based on stimulation of sweat by pilocarpine iontophoresis and collection of the sweat in a disposable plastic device containing a water soluble colorant. Consequently, the gel components employed for sweat stimulation were identified in human sweat either by direct analysis or after sample hydrolysis. Therefore, it would be interesting to reduce the presence of these interferents in collected sweat by implementation of sample

preparation steps. For this reason, the possibility of including centrifugal SPE using C18 or hydrophilic Micro SpinColumn™ was studied. The protocol based on hydrophilic SpinColumn™ was proved by analysis of both the non-retained fraction and the eluate using the HILIC chromatographic mode in both polarities. The resulting chromatograms, shown in Supplementary Fig. 3, were quite similar to those obtained by direct analysis in the case of the eluate and gave no information in the case of the non-retained fraction. Thus, no improvement in terms of metabolite coverage was observed using this SPE protocol.

Concerning sweat preparation by C18 SpinColumn™, analysis of the non-retained metabolites and those in the eluate was independently carried out by both chromatographic columns (C18 and HILIC). Fig. 3 shows the chromatograms obtained in both ionization modes from the non-retained fraction (non-retained metabolites and those washed out in the cleaning step) analyzed using the C18 (Fig. 3a) or HILIC analytical column (Fig. 3b). The chromatograms reveal the degree of efficiency of this SPE step for pilocarpine removal, while methylparaben and propylparaben were completely removed from the sample. The metabolites in the eluate —use of 50% (v/v) acetonitrile for elution—, with a higher non polar character, were analyzed by the chromatographic C18 column for optimum retention. Supplementary Fig. 4 illustrates the chromatograms provided by both ionization modes, in which intense peaks identified as pilocarpine, methylparaben and propylparaben appear. Other exogenous metabolites such as caffeine and theophylline were also identified in the eluate by positive ionization mode. Therefore, this strategy could add interest for sweat cleaning by removal of undesired exogenous metabolites.

Fig. 4a and 4b shows the Venn diagrams comparing the detection coverage in sweat analysis by the C18 analytical column using two different approaches: direct analysis with only dilution, and centrifugal SPE with C18 SpinColumn™ (non-retained and eluted fractions). In negative ionization mode, the most remarkable aspect is the detection of 78 molecular entities in the non retained fraction after SPE versus 29 entities detected by direct analysis. Pilocarpine, methylparaben and propylparaben were only detected in the analysis

after sweat dilution and in the SPE eluate. This effect was similarly observed in positive mode since detection of 69 molecular entities by centrifugal SPE (which included 91% of the entities detected in diluted human sweat) was possible.

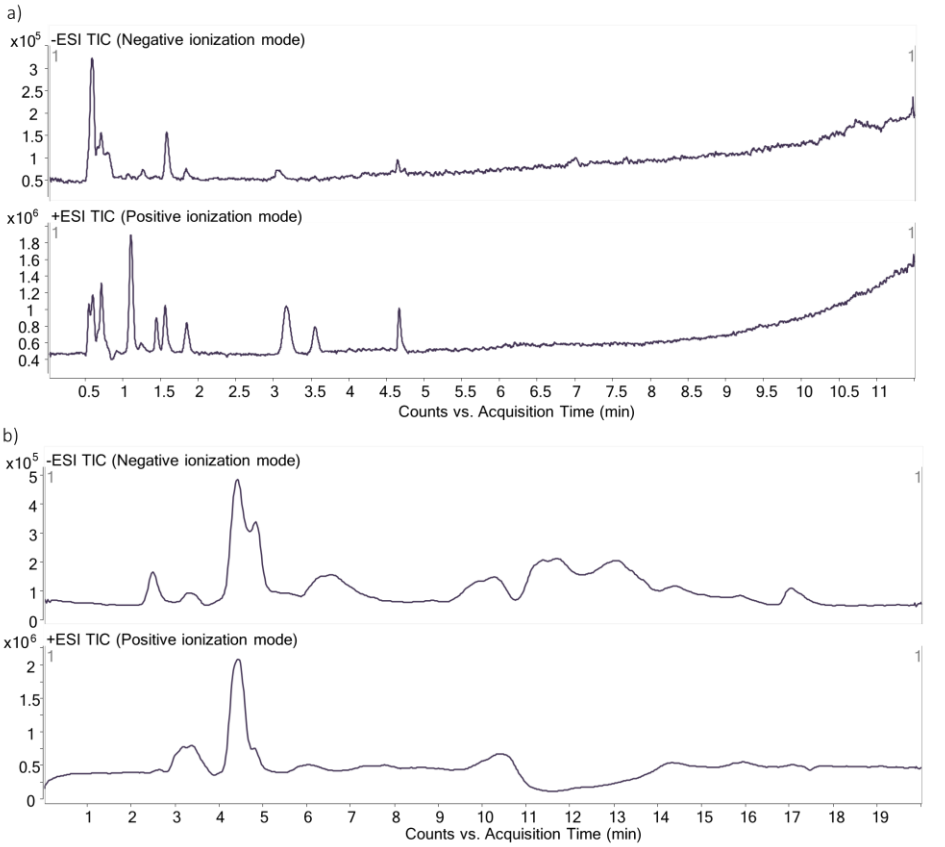


Fig. 3. BPCs obtained in positive and negative ionization modes for the non-retained fraction after centrifugal SPE of human sweat with C18 SpinColumn™ using both chromatographic modes: RP-C18 (a) and HILIC (b).

Fig. 4c and 4d shows the Venn diagrams that compare the detection coverage in the analysis of human sweat by the HILIC analytical column with sample dilution and after centrifugal SPE with the C18 SpinColumn™ (eluate). In this case, more differences were observed between direct analysis and after SPE by the hydrophilic sorbent both in positive and negative ionization modes. In

fact, approximately 40% molecular features detected by analysis of diluted sweat were exclusive of this approach.

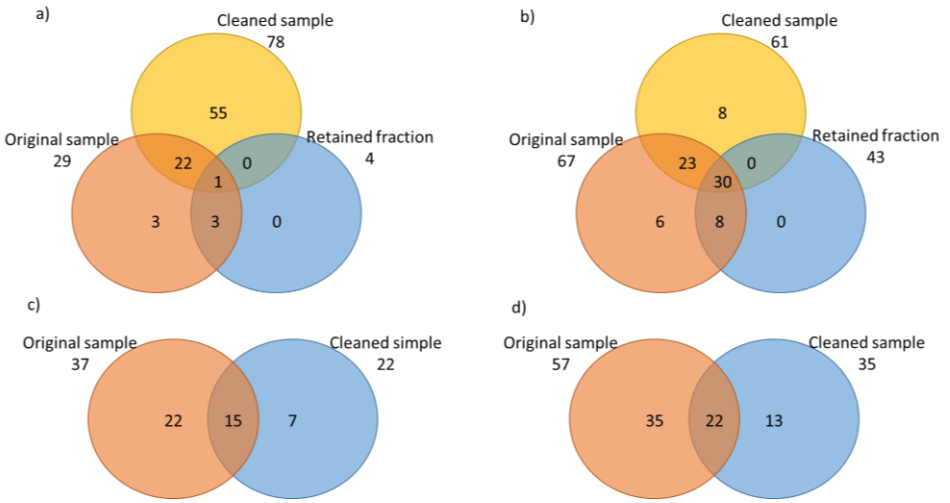


Fig. 4. Venn diagrams comparing the detection coverage in direct sweat analysis and after centrigual SPE using C18 SpinColumn™ (retained and non-retained fractions) using RP-C18 chromatographic mode in negative (a) and positive (b) ionization modes, and using HILIC mode in negative (c) and positive (d) ionization modes.

Attending to these results, the C18 centrifugal SPE approach is appropriate as cleaning strategy by analysis of the non retained fraction since the main interferents associated to the sampling device and exogenous metabolites are retained and subsequently eluted after collection of the non retained target analytes.

3.5. Identification of potential entities detected by MS/MS

Few studies on metabolic composition of sweat have been carried out since the systematic investigations of sweat composition started in the last part of the past century [17]. A recent research by Kutyshenko et al. provided a list of 43 compounds identified by NMR in human sweat including amino acids, sugars and carboxylic acids, among others [12]. However, NMR offers limited sensitivity as compared to MS; therefore, identification of sweat metabolites by MS could increase the metabolomics coverage.

In this research, 41 sweat compounds were identified and confirmed including amino acids, sugars, dicarboxilic acids, fatty acids and derivatives as well as exogenous compounds. Only compounds with high quality of MS/MS spectra at different collision energies were considered in this step. Identification was supported on mass accuracy of precursor ions and that of the most representative fragments, searching for them in the METLIN database. Table 1 lists the 43 identified metabolites, which are classified by families. Information on identification in terms of MS precursor ion and the two most representative product ions is also included. As most of the compounds are detected in both positive and negative ionization polarities, only the ionization mode with the highest sensitivity is included. In all cases, the mass accuracy error, expressed as ppm for precursor ions, ranged from 0 to 4 ppm. The retention time for each compound in both tested chromatographic systems is also included. As exposed above, chromatographic resolution by HILIC methods is better than that by C18 methods as the latter led to coelution of many metabolites in the first part of the chromatogram.

Comparison of the relative intensities for the identified metabolites by the different sample treatments revealed that sample clean up by C18 SpinColumn™ provided the best results in terms of efficiency for most families. Furthermore, most exogenous compounds are almost completely removed from the sample when this strategy is adopted. On the other hand, fatty acids and derivatives were better detected after sample preparation by the hydrophilic SpinColumn™.

The variety of metabolites identified in sweat emphasizes the interest of sweat to be assessed for clinical analysis. Most of the identified metabolites were amino acids and dicarboxylic acids, which are representative of crucial biochemical pathways that could be related to pathologies or dysfunctions. Thus, 19 amino acids, including essential and non essential amino acids, were identified in sweat by combining positive and negative ionization modes. Two amino acids derivatives such as tyramine (tyrosine derivative) and pyroglutamic acid

(glutamic acid derivative) and a dipeptide such as prolylhydroxyproline were unequivocally identified in sweat.

Table 1. List of metabolites identified in human sweat classified by families. Retention time in both chromatographic modes are included as well as the MS identification information (precursor and product ions) and the error (expressed in ppm) of the precursor ion detected.

Family	Metabolite	MS spectra (m/z) [adduct]	MSMS spectra (m/z)	RT (min)	
				C18	HILIC
Amino acids	Arginine	175.1188 $[M+H]^+$	130.0969 116.0709	0.67	-
	Aspartic acid	132.0302 $[M-H]^-$	88.0434 71.0134	0.70	17.64
	Carnitine	116.1124 $[M+H]^+$	103.0391 60.0804	0.67	13.62
	Citrulline	176.1029 $[M+H]^+$	159.0767 113.0702	0.72	15.93
	Glutamate	148.0601 $[M+H]^+$	130.0485 84.0454	0.72	17.26
	Glutamine	147.0759 $[M+H]^+$	130.0504 84.0454	0.69	-
	Histidine	156.0764 $[M+H]^+$	110.0712	0.68	19.19
	Leucine/Isoleucine	132.1017 $[M+H]^+$	86.0966 44.0499	1.43	7.00
				1.61	6.45
	Methionine	150.0581 $[M+H]^+$	133.0310 104.0527	1.14	-
	Ornithine	131.0826 $[M-H]^-$	83.0606	0.72	15.91
	Phenylalanine	166.0860 $[M+H]^+$	120.0806 103.0541	3.55	5.92
	Proline	116.0706 $[M+H]^+$	70.0655	0.78	9.25
	Serine	104.0353 $[M-H]^-$	74.0248	0.67	14.32
	Taurine	124.0074 $[M-H]^-$	79.9565	0.71	7.93
	Threonine	120.0655 $[M+H]^+$	102.0548 74.0606	0.70	17.76
	Tryptophan	205.0971 $[M+H]^+$	188.0705 159.0915	4.67	6.34
	Tyrosine	180.0666 $[M-H]^-$	163.0404 72.0097	1.80	8.34
	Valine	118.0862 $[M+H]^+$	72.0812	0.92	7.56

Amino acid derivatives	Pyroglutamic acid	128.0353 $[M-H]^-$	82.0292	1.56	10.25
	Tyramine	120.0806 $[M+H-H2O]^+$	103.0542 77.0393	3.53	6.10
	Prolyl hydroxyproline	227.1041 $[M-H]^-$	112.0398	4.68	6.42
Amino acid breakdown product	Urocanic acid	137.0357 $[M-H]^-$	93.0459 66.0342	1.16	7.76
Metabolites involved in purine nucleosides degradation	Inosine	267.0728 $[M-H]^-$	135.0310	-	4.99
	Uric acid	167.0211 $[M-H]^-$	124.0157	1.54	10.32
Dicarboxylic acids	Azelaic acid	187.0976 $[M-H]^-$	125.0970 97.4887	6.96	4.22
	Butanedioic acid	117.0191 $[M-H]^-$ 163.0250 $[M+FA-H]^-$	71.0139 101.0246 73.0302	0.91	6.52
	Sebacic acid	201.1134 $[M-H]^-$	183.1038 139.1131	7.62	-
	Suberic acid	173.0824 $[M-H]^-$	111.0809 83.0503	6.21	-
Exogenous compounds	Caffeine	195.0879 $[M+H]^+$	138.0655 110.0708	5.31	2.87
	Methylparaben	151.0413 $[M-H]^-$	136.0172 92.0278	7.56	2.50
	Pilocarpic acid	225.1246 $[M-H]^-$	163.1218	2.93	10.63
	Pilocarpine	209.1279 $[M+H]^+$	163.1227 95.0608	3.35	3.01
	Propylparaben	179.0714 $[M-H]^-$	136.0153 92.0262	9.33	2.51
	Theophylline	181.0712 $[M+H]^+$	124.0498 69.0444	4.76	3.18
Nutrients	Choline	104.1070 $[M+H]^+$	60.0819 45.0340	0.64	6.36
	Coumaric acid	165.0554 $[M+H]^+$	119.0499	1.79	8.32
Fatty acids derivatives	MG (18:0)	359.3148 $[M+H]^+$	341.3042 85.1008	-	2.48
	MG (16:0)	331.2831 $[M+H]^+$	313.2711 123.1155	10.44	2.50
	MG (22:2)	431.3127 $[M+Na-2H]^-$	89.0247	10.12	2.60
Sphingolipids	C16 Sphinganine	274.2742 $[M+H]^+$	257.2480	8.76	-
Sugars	Maltotriose	503.1636 $[M-H]^-$	161.0463 341.1094	0.79	16.58
	Tetrasaccharide	689.2113 $[M+Na]^+$	325.1123 487.1625	0.81	18.78

Concerning dicarboxylic acids, azelaic acid, butanedioic acid, sebacic acid and suberic acid were confirmed in negative ionization mode. The identification of four dicarboxylic acids reveals the importance of this group of endogenous metabolites in human sweat.

Two metabolites involved in the degradation of purine nucleosides such as inosine and uric acid were also detected in human sweat. Therefore, the activity of this pathway is also reflected in this non invasive sample. Urocanic acid, a breakdown product of histidine, was identified as excreted metabolite in human sweat. This compound is accumulated in the epidermis where it acts as a UV protector and as immunoregulator. One other low molecular weight metabolite identified in sweat was choline, which can be either ingested as free form in the diet or produced by degradation of glycerophosphocholines. Previous studies have described the presence of this metabolite in sweat [12,13].

Finally, it is worth mentioning the identification of one sphingolipid and three fatty acid derivatives with non polar character —viz. C16 sphinganine and three monoacylglycerols such as MG(18:0), MG(16:0) and MG(22:2)—, which emphasize the presence of non polar metabolites in sweat, despite the aqueous nature of this biofluid.

Among the 43 compounds identified in this research using LC–QTOF some of them (viz. coumaric acid, uric acid, dicarboxylic acids, carnitine, taurine, tyramine, tyrosine and fatty acid related compounds) were not identified by a previous study based on NMR [12]. In addition, some of the identified metabolites had been evaluated as potential biomarkers in other biofluids [18-20]; aspect that may be of interest to study sweat as potential biofluid for biomarker searching.

4. Conclusions

Different sample preparation protocols as well as different chromatographic modes have been compared for global metabolomics profiling analysis

of sweat. Although sweat is scarcely used in metabolomics studies, its potential for discrimination of metabolic profiles correlated with different pathologies would make this biofluid interesting to explore. Among the tested protocols, sweat hydrolysis would not be recommended for metabolite profiling since the sample integrity is altered by detection of numerous artifacts. Direct analysis after dilution was a suited option to obtain a representative snapshot of sweat metabolome. On the other hand, the clean up step with C18 SpinColumn SPE cartridges improves sensitivity and reduces the number of exogenous compounds. Only for fatty acids the clean up step should be developed by hydrophilic SpinColumn SPE cartridges.

Identification of a high number of sweat components from a wide variety of families demonstrated that LC–QTOF MS/MS is a good strategy to analyze this biofluid. It should be mentioned that almost 50% of the identified compounds were amino acids, being the compounds more abundant in sweat.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2012-37428). F. Priego-Capote is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks the MICINN for an FPU scholarship (AP2009-0499). Reina Sofía Hospital (Córdoba, Spain) is also grateful for donors of sweat samples.

References

[1] S.C. Way, A. Memmesheimer, Arch. Derm. Syphilol. 41(6) (1940) 1086.

[2] B. Schittek, R. Hipfel, B. Sauer, J. Bauer, H. Kalbacher, S. Stevanovic, M. Schirle, K. Schroeder, N. Blin, F. Meier, G. Rassner, C. Garbe, Nat. Immunol. 2 (2001) 1133.

[3] L.E. Gibson, R.E. Cooke, Pediatrics 23(3) (1959) 545.

[4] A. Lynch, Thesis: Development and characterisation of solid-state ion-selective electrodes, and their application to sweat analysis for cystic fibrosis diagnosis, School of Chemical Sciences, Dublin City University, 2010.

[5] J.M. Kirk, M. Keston, I. McIntosh, S. Essa, Ann. Clin. Biochem. 29 (1992) 145.

[6] N. De Giovanni, N. Fucci, Curr. Med. Chem. 20(4) (2013) 545.

[7] P. Kintz, A. Tracqui, P. Mangin, J. Anal. Toxicol. 20 (1996) 393.

[8] J.R. Taylor, I.D. Watson, F.J. Tames, D. Lowe, Addiction 93 (1998) 847.

[9] N. Uemura, R.P. Nath, M.R. Harkey, G.L. Henderson, J.Mendelson, R.T. Jones, J. Anal. Toxicol. 28 (2004) 253.

[10] A. Mena-Bravo, M.D. Luque de Castro, J. Pharma. Biomed. Analysis, doi: 10.1016/j.jpba.2013.10.048.

[11] B.M.R. Appenzeller, C. Schummer, S. Boura Rodrigues, R. Wennig, J. Chromatogr. B 852 (2007) 333.

[12] V.P. Kutyshenko, M. Molchanov, P. Beskaravayny, V.N. Uversky, M.A. Timchenko, PLoS ONE 6(12) (2011) e28824.

[13] N. Papanas, A.J.M. Boulton, R.A. Malik, C. Manes, O. Schnell, V. Spallone, N. Tentolouris, S. Tesfaye, P. Valensi, D. Ziegler, P. Kempler, Diabet. Med. 30 (2013) 525.

[14] M. Beauchamp, L.C. Lands, Pediatric Pulmon.39 (2005) 507.

[15] N.A.S. Taylor, C.A. Machado-Moreira, Extrem. Physiol. Med. 2 (2013) 4.

[16] M.M. Raiszadeh, M.M. Ross, P.S. Russo, M.A. Schaepper, W. Zhou, J. Deng, D.Ng, A. Dickson, C. Dickson, M. Strom, C. Osorio, T. Soeprono, J.D.Wulfkuhle, E.F. Petricoin, L.A. Liotta, W.M. Kirsch, J. Proteome Res. 11 (2012) 2127.

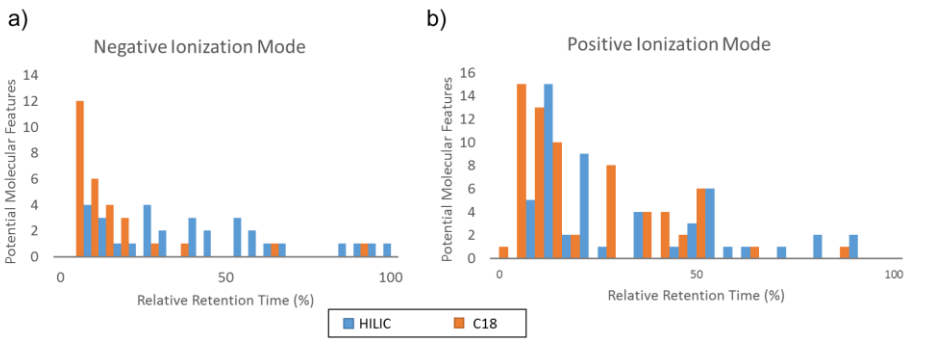
[17] M. Harker, H. Coulson, I. Fairweather, D. Taylor, C.A. Daykin, Metabolomics 2 (2006) 105.

[18] I.M. El Agouza, S.S. Eissa, M.M. El Houseini, D.E. El-Nashar, O.M. abd El Hameed, Angiogenesis 14(3) (2011) 321.

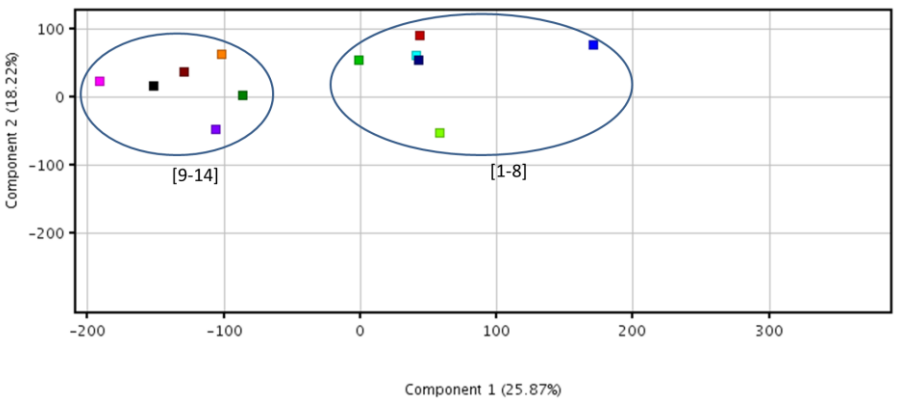
[19] A.C.M. Gagliardi, M.H. Miname, R.D. Santos, Atherosclerosis 202(1) (2009) 11.

[20] M.G.M. de Sain-van der Velden, P. Rinaldo, B. Elvers, M. Henderson, J.H. Walter, B.H.C.M.T. Prinsen, N.M. Verhoeven-Duif, T.J. de Koning, P. van Hasselt, JIMD Reports 6 (2012) 95.

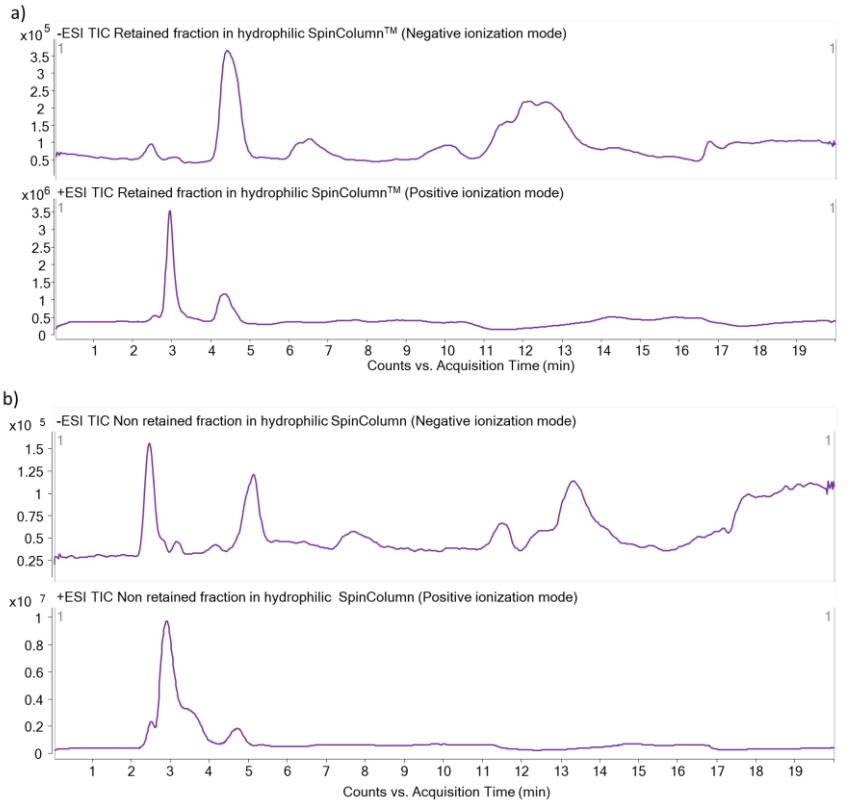
Supplementary material



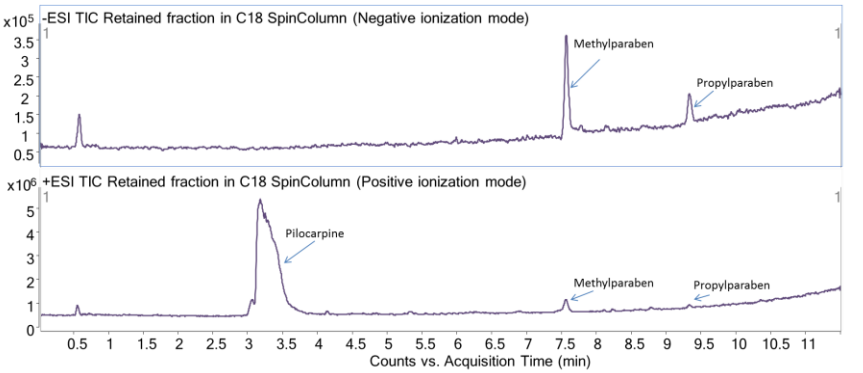
Supplementary Fig. 1. Distribution of the retention times for the molecular features detected by LC–MS/MS analysis of human sweat with each chromatographic mode and ionization mode.



Supplementary Fig. 2. Principal Component Analysis (PCA) as quality control study to evaluate instrumental precision by analysis of human sweat pool using the signal of methylparaben as normalization strategy.



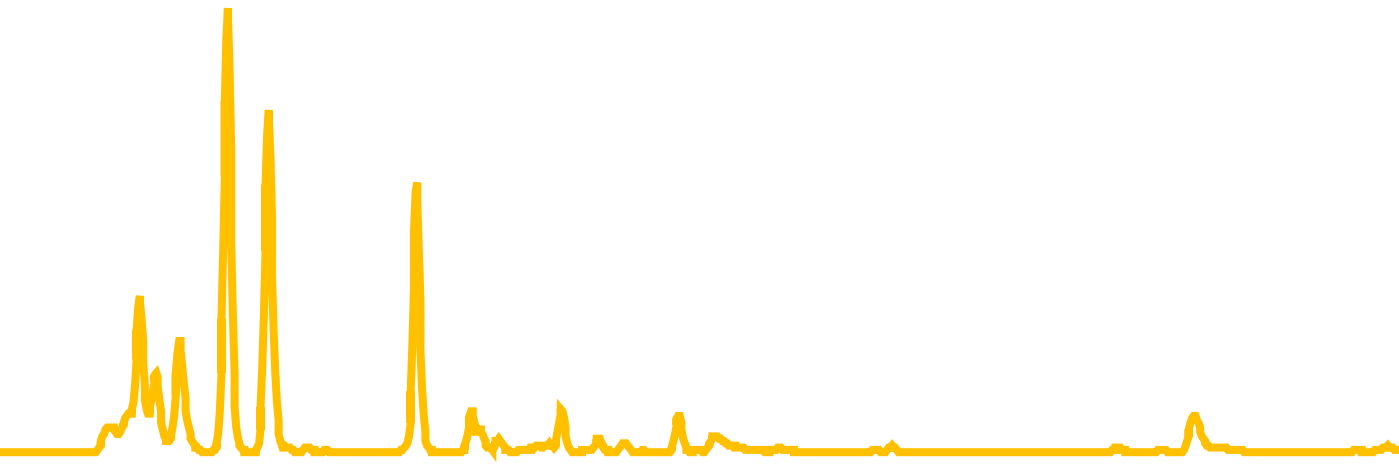
Supplementary Fig. 3. BPCs obtained after centrigugal SPE of human sweat with hydrophilic SpinColumn™ by analysis of the retained fraction (a) and non-retained fraction (b).



Supplementary Fig. 4. BPCs obtained by analysis of the retained fraction in centrifugal SPE of human sweat with C18 SpinColumn™ in positive (a) and negative (b) ionization modes.

Chapter 2:

Enhancing detection coverage in global metabolomics profiling by automated solid phase extraction on-line coupled to LC–MS/MS



Send to

Journal of Chromatography A

Enhancing detection coverage in global metabolomics profiling by automated solid-phase extraction on-line coupled to LC–MS/MS

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b]*, *M. D. Luque de Castro*[a,b]

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

Enhancing detection coverage in global metabolomics profiling by automated solid-phase extraction on-line coupled to LC–MS/MS

M. Calderón-Santiago, F. Priego-Capote, M. D. Luque de Castro*

Abstract

One of the main shortcomings of global metabolomic profiling is a limited detection coverage that, however, can be improved by a combination of analytical techniques such as NMR spectroscopy, LC–MS or GC–MS. An automated approach was configured by on-line coupling of solid-phase extraction (SPE) to LC–MS. The approach was applied to the metabolomic analysis of human serum based on sample fractionation by this SPE–LC–MS coupling. With this purpose, sorbents acting either via non-mixed or mixed retention mechanisms were assessed for their ability to provide selective fractions of serum metabolites. A combination of the results provided by the SPE protocols for the different sorbents allowed 3445 molecular entities to be detected —more than 73% in the fractions eluted from at least two sorbents. Also, more than 81% of this detection coverage was obtained by a combination of analyses carried out in the eluates from four sorbents: polydivinylbenzene resin, C18 sorbent, and anionic and cationic mixed-mode polymeric sorbents. Also, pairs of SPE protocols were serially conducted in automated SPE–SPE–LC–MS configurations to boost their selectivity in single analyses and hence maximize throughput, which can be especially useful with scarce or highly valuable samples. The serial configurations afforded single-analysis detection coverages of 79.5–99.7% for molecular features via two separate analyses with the same sorbents.

1. Introduction

The primary purpose of metabolomics is to analyze the complement of low-molecular weight compounds present in a biological fluid, cell or organism under a given set of physiological conditions or different perturbations. Metabolomics essentially relies on three major strategies depending on whether analyses are to be comprehensive, qualitative and/or quantitative [1]. So-called "targeted analyses" involve the qualitative and quantitative study of one or, more frequently, a small group of chemically similar metabolites. On the other hand, metabolomics fingerprinting uses a global screening approach to classify samples in terms of metabolite patterns (fingerprints) that change in response to a specific factor. Finally, global metabolomics profiling or "untargeted analysis" involves the 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. The greatest challenge of metabolomics is maximizing detection and accurate identification of thousands of metabolites, which is currently a utopian goal in dealing with complex organisms. For instance, viral and bacterial metabolomes contain around a thousand or even fewer metabolites [2,3], whereas the human metabolome consists of approximately 7900 metabolites [4] and plant metabolomes can easily contain up to 100 000–200 000 metabolites [5,6]. The determination ability of platforms for metabolomics analysis can be measured in terms of "metabolomics coverage", that is, the number of metabolites an individual platform such as GC–MS/MS, LC–MS/MS, CE–MS/MS or NMR spectroscopy, or a combination thereof, can detect.

The main limitation of existing metabolomics profiling methodologies is a result of the high complexity of the metabolome, which includes a wide range of compounds with different physico–chemical properties at concentrations spanning several orders of magnitude [7,8]. Specifically, metabolite levels can range from a few picomoles to a few millimoles per litre, which complicates the detection of low-concentrated metabolites in the presence of metabolites existing at high concentration levels.

The most effective strategy for the comprehensive analysis of metabolomes currently available involves integrating information obtained with different analytical platforms (mainly NMR spectroscopy [9,10] and MS coupled to a separation technique such as GC, LC or CE [11–13]) or using direct infusion [14,15]. One other way of maximizing metabolomic coverage is by optimizing existing fractionation protocols for sample preparation [16,17]. These protocols are intended to enable the analysis of metabolites according to physico-chemical properties such as polarity, acid–base character, affinity or molecular size. Sample preparation in metabolomics usually includes liquid–liquid extraction (LLE) [18–20] or solid–liquid extraction with different solvents with special emphasis on solid-phase extraction (SPE) with selective sorbents [21,22]. However, the poor automatability of these strategies makes protocols time-consuming. For example, the metabolomics profiling of *Escherichia coli* has been addressed by integrating the analysis of extracts obtained with different solvent compositions [23]. This strategy, however, cannot be automated, so it is impractical for analyzing large numbers of samples. Unlike liquid–liquid extraction and solid–liquid extraction, SPE can be easily automated to extract metabolites from biological fluids with a wide variety of sorbents capable of retaining compounds with different properties [24-26]. Integrative SPE with selective sorbents for subsequent LC–MS analysis can be expected to improve metabolomics coverage.

In this work, an automated platform was used for integrative SPE coupled on-line to LC–MS/MS analysis of human serum in high resolution mode with a view to enhancing metabolomics coverage in relation to conventional protocols currently used to characterize the serum metabolome. In addition, the proposed platform allows complementary SPE protocols to be combined in purposely constructed SPE–SPE–LC–MS configurations for more comprehensive analysis of biofluids. The ultimate aim was to increase the number of metabolites detected in order to acquire more biological information for subsequent processing.

2. Experimental section

2.1. Chemicals

MS-grade acetonitrile (ACN), formic acid, ammonia and ammonium formate were purchased from Scharlab (Barcelona, Spain) and used to prepare chromatographic mobile phases and the solutions used in the SPE protocol. Deionized water (18 mΩ · cm) from a Millipore Milli-Q water purification system was used throughout.

2.2. Instruments and apparatus

An Agilent 1200 Series LC system coupled to an Agilent Accurate-Mass QTOF hybrid mass spectrometer equipped with a dual source for electrospray ionization (ESI) (Santa Clara, CA, USA) was used to construct the analytical platform (Fig. 1). The chromatographic eluate was monitored in the accurate mass mode.

SPE was done in a Symbiosis Pharma automated workstation from Spark Holland (Emmen, The Netherlands) equipped with a Reliance autosampler, also from Spark Holland, and a refrigerated stacker sample compartment. The autosampler was furnished with a 0.2 mL sample loop. The workstation contained two high-pressure syringe dispensers (HPD) for solvent delivery and four 6-port valves connected with Peek tube of 0.25 mm i.d. (VICI, Houston, Texas, USA), which was used for all system connections. The SPE workstation was equipped with a unit for cartridge exchange —automatic cartridge exchanger (ACE)— holding up to 100 cartridges which were clamped in the loop of two of the valves (Clamp I and Clamp II) for each analysis. A 1 mL × 1.0 mm i.d. stainless tube from Análisis Vínicos (Tomelloso, Spain) was used to connect the two valves of the SPE workstation (*viz.* Valve 2 and Clamp II). This configuration allowed the eluate from the first cartridge (Clamp I) to be conditioned for a second SPE step (Clamp II), if required, by mixing with other solution prior to loading in a complementary second cartridge. Cartridges (10 × 2 mm) packed with the following sorbents were tested: Hysphere C2 (silica-based ethyl phase),

Hysphere C8 (EC) (end-capped silica-based octyl phase), Hysphere C18 (silica-based octadecyl phase), Hysphere Resin GP (polymeric polydivinylbenzene phase), Hysphere Resin SH (strong hydrophobic-modified polystyrene–divinylbenzene), Hysphere MM anion (mixed-mode anionic), Hysphere MM cation (mixed-mode cationic, Spark Holland) and Oasis HLB (hydrophilic–lipophilic balance, Waters, Milford, MA, USA). The fourth valve, located behind the column compartment, was used to select the chromatographic fraction to be monitored in LC–MS/MS..

2.3. Blood collection and serum isolation

Venous blood was collected in evacuated sterile serum tubes (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) containing no additives and incubated at room temperature for 10 min to facilitate coagulation. Then, the tubes were centrifuged at 2000 *g* at 4 °C for 15 min to isolate the serum fraction (processing within 2 h after collection). A serum pool prepared by mixing different serum samples and stored at –80 °C until analysis was used in all optimization tests.

2.4. LC–TOF/MS analysis

Chromatographic separation was performed at 25 °C, using a C18 reversed phase analytical column (Mediterranean, 100 mm × 4.6 mm i.d., 3 μm particle size) 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 or 5 mM ammonium formate (pH 6.1) when using cationic cartridges. The LC pump was programmed to operate at a flow rate of 0.8 mL/min in accordance with the following gradient elution sequence: 3% phase B for 1 min; phase B ramp from 3 to 100% from min 1 to min 25; and hold at 100% for 3 min.

The electrospray ionization source was operated in the 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 and 325 °C, respectively. The

instrument was calibrated and tuned as recommended by the manufacturer. Data were collected in the centroid mode, using a rate of 1.0 spectrum per second in the extended dynamic range mode (2 GHz). Accurate mass spectra were acquired over the *m/z* range 100–1100. The instrument gave typical resolution 15 000 FWHM (Full Width at Half Maximum) at *m/z* 118.086255 and 30 000 FWHM at *m/z* 922.009798. Mass accuracy in recorded ions was assured by continuous internal calibration via the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated *hexakis*(1H, 1H, 3H-tetrafluoropropoxy)phosphazine or HP-921] in the positive ionization mode; and those at *m/z* 112.9856 (trifluoroacetate anion) and *m/z* 1033.9881 (HP-921) in the negative mode.

2.5. Single cartridge automatic SPE configuration and protocol

The retention capabilities of the different cartridges were assessed separately, using a single cartridge configuration (Fig. 1). For this purpose, four standard SPE protocols (see Table 1) were programmed for application in terms of the characteristics of each sorbent, namely: two reversed phase protocols differing only in the loading solution, a cationic protocol and an anionic protocol. In all cases, retained metabolites were eluted by using a suitable mobile phase in the programmed gradient of the chromatographic method. Variability in the process was assessed by using three replicates per sorbent (see Supplementary Fig. 1).

Table 1. Solid phase extraction protocols used with the different sorbents.

Step	Reversed phase protocol 1	Reversed phase protocol 2	Cationic protocol	Anionic protocol
Solvation	1 mL ACN	1 mL ACN	1 mL ACN	1 mL ACN
Equilibration	1 mL water	1 mL 20% ACN	1 mL 20% ACN 1% formic acid	1 mL 20% ACN 2% ammonia
Sample loading	1 mL water	1 mL 40% ACN	1 mL 20% ACN 1% formic acid	1 mL 20% ACN 2% ammonia
Elution	Mobile phase with formic acid	Mobile phase with formic acid	Mobile phase with ammonium formate	Mobile phase with formic acid

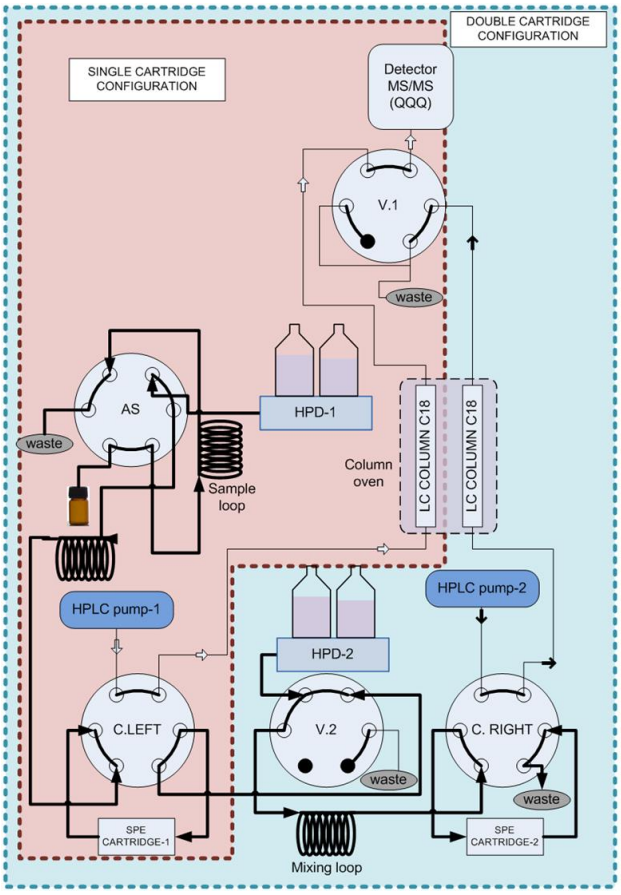


Figure 1. Instrumental configurations for single- and double-cartridge SPE on-line coupled to LC–TOF/MS including an autosampler (AS), two high-pressure dispensers (HPDs), two clamps and five two-position valves (one with a T-rotor).

2.6. Double cartridge automatic SPE configuration and protocol

The double cartridge configuration used to combine complementary cartridges is also shown in Fig 1. The SPE protocols were similar (see Table 1) and differed only in equilibration or in the solution loaded into the second cartridge. HPD 2 was used to directly pump an appropriate solution to equilibrate the second cartridge or for mixing with the eluate from the first cartridge. The eluates from both cartridges were sequentially led to the chromatographic column for independent, consecutive analysis.

2.7. Data processing and statistical analysis

MassHunter Workstation software (version 3.01 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) was used to process all data obtained by LC–TOF/MS in the full scan MS mode. Processing of raw data files started by extracting potential molecular features (MFs) with the applicable algorithm included in the software. The extraction algorithm considered all ions exceeding 5000 counts (above background noise) with a single charge state. In addition, the isotopic distribution for a valid feature had to be defined by two or more ions (with peak spacing tolerance of 0.0025 *m/z*, plus ±7.0 ppm). Adduct formation in the positive (+Na, +K, $+NH_4$) and negative ionization mode (+Cl, +HCOO), and neutral losses by dehydration and the loss of phosphate, glucuronide or methyl groups, were also included to identify features of the same metabolite. Thus, ions with identical elution profiles and related *m/z* values (representing different adducts or isotopes of the same compound) were extracted as entities characterized by their retention time (RT), intensity at the apex of each chromatographic peak and accurate mass. In this way, raw data files were created in compound exchange format (.cef files) for each sample and exported into the Mass Profiler Professional (MPP) software package (version 2.0, Agilent Technologies, Santa Clara, CA, USA) for further processing. In a subsequent step, the data were preprocessed by aligning retention times and *m/z* values across the data matrix, using a tolerance window for retention time and mass accuracy of 0.5 min and 7 ppm, respectively. Venn diagrams were then constructed to compare the number of entities found with each cartridge or configuration.

3. Results and discussion

As stated before, the selectivity of the retention mechanisms for SPE sorbents can provide an interesting analytical tool for assisting fractionation of complex metabolomes.

This research was entirely conducted with a pool serum prepared from several healthy donors to ensure comparability of tests, all of which were performed in triplicate. Also, all chromatographic separations were done by using a C18 phase column in the reversed gradient mode. HILIC chromatographic columns were not included in this study since the elution step was carried out in gradient mode for all tested sorbents. Thus, HILIC-gradient methods were not compatible with this elution mode. Retention capabilities were optimized for comparison of the sorbents based on non-mixed retention mechanisms first and those relying on mixed mechanisms then. Sorbents providing wider coverage were also assayed, using the above-described serial configurations

3.1. Retention capabilities of SPE reversed phase sorbents based on non-mixed mechanisms

The materials acting via non-mixed retention mechanisms studied here included C18, C8 and C2, in increasing polarity —a function of the hydrocarbon chain length. The retention ability of each material was assessed by inserting on-line the eluted fraction from the column into the LC–TOF/MS system to determine the number of metabolites retained in the sorbent. The results are shown as a Venn diagram in Fig. 2.A. As can be seen, the C2, C8 and C18 sorbents allowed a quite similar number of molecular features to be determined (*viz*. 1421, 1524 and 1593, respectively; and up to 2178 entities in combination). Also, a minimum of 1459 common molecular features, which represented 67.0% of the total detection potential, were detected in at least two of the tests. Based on these results, the C18 sorbent was seemingly the best for obtaining a representative view of the metabolome of human serum judging by its detection capabilities. However, C8 and C2 should not be discarded as they enabled the detection of a substantial number of molecular entities not covered by C18. Supplementary Fig. 2.A shows the base peak chromatograms (BPCs) in the positive ionization mode obtained by LC–TOF/MS analysis of the eluates from the three SPE sorbents. The analysis of the fractions extracted with C18, C8 and C2 gave BPCs with wide common chromatographic zones and a few different zones. The differences between the results obtained in the negative and positive ionization modes were

quite similar (see BPCs in Supplementary Fig. 2.B); also, the eluates from the C8 and C2 sorbents exhibited decreased retention efficiency relative to C18. Figure 2.B shows the extracted ion chromatograms (EICs) for *m/z* 518.3217 corresponding to LysoPC(18:3), which represents to a characteristic family of serum metabolites detected in the eluted fraction of the three sorbents. As can be seen, the retention efficiency of LysoPC(18:3) in C18 sorbent was clearly higher than that observed for the other two tested materials.

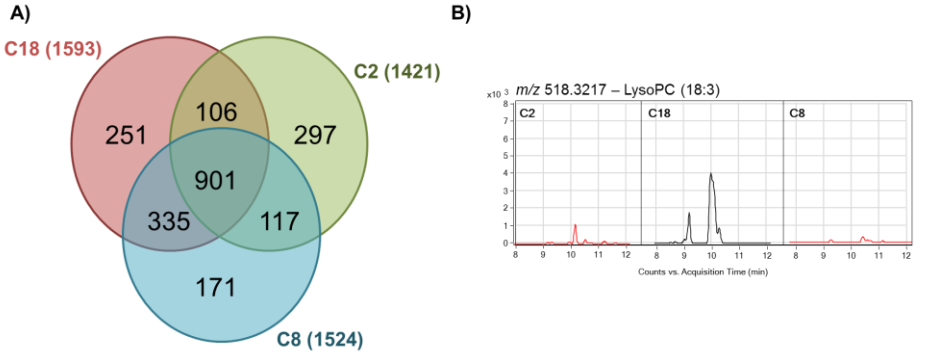


Figure 2. A) Venn diagram comparing molecular features (MFs) determined with a single-cartridge SPE–LC–TOF/MS configuration and sorbents based on non-mixed retention mechanisms (C18, C8 and C2) in the positive ionization mode. **B)** Extracted ion chromatogram (EIC) for precursor ion at *m/z* 518.3217 corresponding to lysoPC (18:3).

3.2. Retention capabilities of SPE sorbents based on mixed mechanisms

Materials simultaneously acting via mixed retention mechanisms are frequently used to expand the applicability of SPE. In this work, we used five types of sorbents based on different retention mechanisms, namely: polymeric resins GP (polymeric polydivinylbenzene phase) and SH (strong hydrophobic-modified polystyrene-divinylbenzene); Oasis HLB (a sorbent with balanced hydrophilic and lipophilic interactions); and two sorbents exhibiting ionic interactions (an anionic and a cationic MM sorbent). These sorbents were tested with the reversed phase protocol previously used for sorbents involving non-mixed retention mechanisms. As can be seen in Fig. 3.A, the eluate from resin GP provided greater coverage of metabolites in human serum than the other

sorbents acting via a combined retention mechanism (*viz.* 1596 entities versus 1215 with resin SH and 1119 with HLB). In fact, 78% of the entities detected in the eluate from resin GP were also detected with HLB and SH. Based on these results resin GP is seemingly especially suitable for obtaining a representative profile of metabolites in human serum under the tested conditions. Supplementary Fig. 2.C shows the BPCs obtained by analyzing the eluates from the resing GP, resin SH and Oasis HLB; as can be seen, there were critical differences consistent with the Venn diagrams. Figure 3.B and C illustrate EICs for *m/z* 203.0526 corresponding to a sodium adduct of a hexose such as glucose, fructose or mannose and *m/z* 132.1019 associated to leucine. Both metabolites are representative examples of high and low concentrated metabolites in serum, detected in the eluted fraction of the three sorbents. The retention efficiency of these compounds in resin GP sorbent was clearly higher than that observed in the other two polymeric materials.

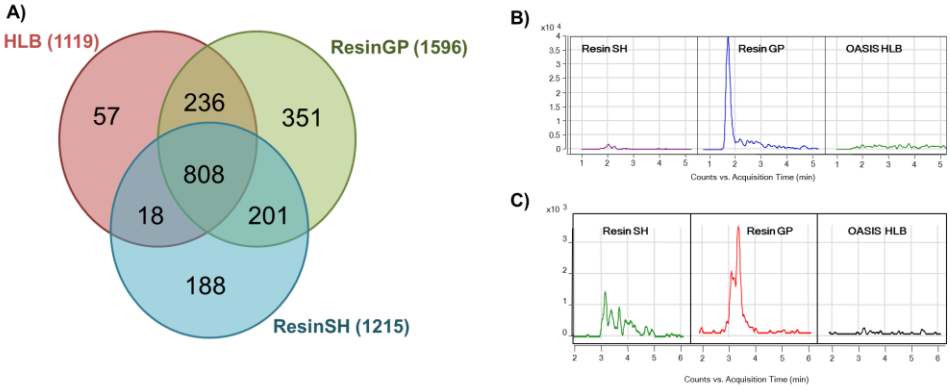


Figure 3. A) Venn diagram comparing molecular features (MFs) determined with a single cartridge SPE–LC–TOF/MS configuration and sorbents based on mixed retention mechanisms (resin GP, resin SH and HLB Oasis). **B)** EIC for precursor ion at *m/z* 203.0526 associated to a sodium adduct of a hexose. **C)** EIC for a precursor ion at *m/z* 132.1019 corresponding to leucine.

Ionic MM sorbents were studied by using a protocol similar to that for the reversed SPE procedures but at a different pH in two key steps: sample loading and elution. In this way, ionic interactions were enhanced via the

formation of formic acid and ammonium formate, which facilitated retention of cationic and anionic compounds in both cartridges (see Table 1). The two cartridges contained the same polymeric support and differed only in their cationic or anionic nature; therefore, their differences in retention capabilities can be exclusively ascribed to ionic interactions and the influence of an acid or base on the loading and elution steps. As can be seen from Fig. 4, a high proportion of molecular features (60% with the anionic MM sorbent and 80% with the cationic MM sorbent) were detected in the eluates from both cartridges. Also, the sorbents proved quite selective: independent experiments with the ionic MM sorbents allowed 316 and 122 features to be uniquely detected. These differences, resulting from the selectivity of the sorbents, reflected in marked overlap between the BPCs for the two sorbents (see Supplementary Fig. 2.D). Figure 4 also shows EICs of two metabolites, thromboxane A2 (*m/z* 353.2322) and 1-methyladenosine (*m/z* 282.1196), which were preferably retained in one of the ionic MM sorbents. Thus, thromboxane A2 was well-retained in the anionic polymeric sorbent while 1-methyladenosine was properly retained in the cationic polymeric sorbent. These examples support the complementary behavior of both SPE cartridges.

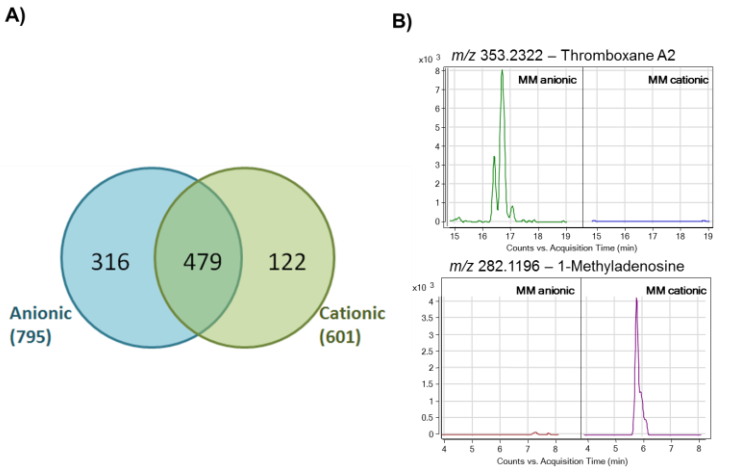


Figure 4. A) Venn diagram comparing molecular features (MFs) determined with a single cartridge SPE–LC–TOF/MS configuration and sorbents based on mixed retention mechanisms (anionic and cationic MM sorbents). **B)** EICs for precursor ions at *m/z* 353.2322 and 282.1196 corresponding to thromboxane A2 and 1-methyladenosine, respectively.

3.3. Detection coverage of SPE–LC–TOF/MS with the studied sorbents

Once assessed for retention, the sorbents were compared in terms of the number of molecular features detected in their eluates. Table 2 lists the number of entities detected by SPE–LC–TOF/MS in each fraction. Aligning the TOF chromatograms via the retention time and mass accuracy afforded the detection of 3445 molecular entities in combination —more than 73% in the eluates from at least two sorbents. Supplementary Fig. 3 shows the extracted ion chromatograms obtained in the analysis of different eluates for three of the metabolites present in human serum. In view of these results, sample fractionation by SPE could be an approach for maximizing the detection capabilities of some analytical methodologies.

Table 2. Molecular features detected in the SPE–LC–TOF/MS analysis of the fractions eluted from each sorbent in the single-cartridge configuration. The total number of unique molecular features detected, and the proportion detected with each individual sorbent, are also shown.

Sorbent	Number of MFs	Detection coverage (%)
C18	1593	46.2
C8	1524	44.2
C2	1421	41.2
Resin GP	1596 (801)*	46.3
Resin SH	1215 (603)*	35.3
Oasis HLB	1119 (403)*	32.5
MM anion	795	23.1
MM cation	601	17.4
Total	3445	

* The number of entities detected with the nonpolar protocol is given in brackets.

Psychogios *et al.* in the framework of the Human Metabolome Database Consortium [27] succeeded in detecting and identifying 3564 metabolites by using a combination of five different methodologies based on LC–MS/MS, NMR, GC–MS, direct infusion–MS and TLC prior to GC–FID. In the research proposed here no overall identification was carried out by tandem mass spectrometry since

the main objective was to show that the selectivity of these sorbents can be exploited jointly in on-line SPE–LC–MS analysis of biological samples to substantially increase the metabolite coverage of this detection technique.

The detection capability of the SPE–LC–TOF/MS approach with combination of sorbents was also assessed by comparison with the analyses of fractions obtained by serum liquid–liquid extraction with a methanol–chloroform system. To this end, a 300 µL aliquot of the serum pool was mixed with an identical volume of a 1:2 chloroform–methanol solution to precipitate proteins. The methanolic and chloroform fractions were separated by centrifugation and analyzed by LC–TOF/MS. As can be seen, the combined analysis of the two fractions enabled detection of 1444 entities, which is below the detection capability of the automated SPE approach. Figure 5 compares the results obtained by the SPE–LC–TOF/MS approach and the precipitation method involving a binary mixture of methanol and chloroform.

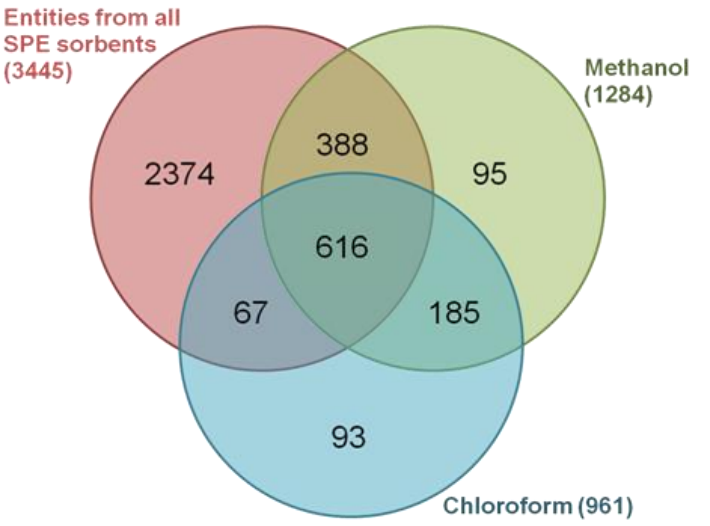


Figure 5. Venn diagram comparing molecular features (MFs) determined by SPE–LC–TOF/MS combining all sorbents tested in comparison with a method based on liquid–liquid extraction with methanol–chloroform mixture and independent analysis of the two fractions.

The potential of sorbents individually exhibiting the highest detection coverage in SPE–LC–TOF/MS was assessed with a view to their joint, complementary use. The target sorbents were C18 and resin GP, which were used with conventional reversed phase protocols, as well as cationic and anionic MM sorbents, which were used with pH-modified protocols. Supplementary Fig. 4 shows a Venn diagram comparing the metabolite coverage achieved by analyzing the respective eluates. As can be seen, only 10 molecular features were shared by the four eluates; this underlines the fractionation capabilities of the selected materials. Using the four eluates in combination allowed the detection of a total of 2812 molecular entities —a much higher figure than that provided by independent analyses with the individual sorbents. Many of the features (62%) were detected in the eluates from at least two sorbents. In addition, the results of these joint analyses were complementary. Thus, 881 unique features were detected in the eluates from the ionic MM sorbents, with 459 common entities detected with the anionic and cationic protocols. On the other hand, 1895 molecular features were exclusively detected with the protocols based on SPE in the reversed mode (C18 and resin GP). Finally, the analyses with C18 and resin GP as sorbents allowed the detection of 1240 common features.

3.4. Serial configurations for improved metabolomics coverage (double cartridge)

Once the complementarity of the fractions obtained with different cartridges was demonstrated, technical configurations based on two serially arranged SPE cartridges were designed to exploit the benefits of orthogonal SPE protocols. This allowed metabolites not retained in the first cartridge to be trapped in the second. As shown earlier, cationic and anionic MM sorbents provided highly selective metabolite profiles by virtue of their ionic interactions. Therefore, the most suitable combinations for the intended purpose should be those using a reversed mode sorbent such as C18 or GP resin with an ionic MM sorbent in order to maximize metabolomic coverage. Figure 6 shows Venn diagrams illustrating the detection capabilities of the double cartridge approach with a single serum aliquot. As can be seen, the number of molecular features

detected in the eluates from the ionic MM sorbents used in the second cartridges was much smaller than that afforded by separate analysis with ionic sorbents in a single cartridge configuration. The orthogonal protocols allowed detection of a substantial number of molecular entities (1884 to 2176). Such high detection ability arises from the supplemental selectivity provided by these compatible SPE protocols. Supplementary Fig. 5 shows one of the orthogonal BPCs obtained by sequential elution from the two cartridges used in one of the configurations. Table 3 compares the detection capabilities of the serial configurations in a single analysis with those of a combination of two separate analyses with the same two sorbents. As can be seen, the detection coverage with the serial SPE system ranged from 79.5 to 99.7% of the molecular features detected with a combination of the two separate analyses.

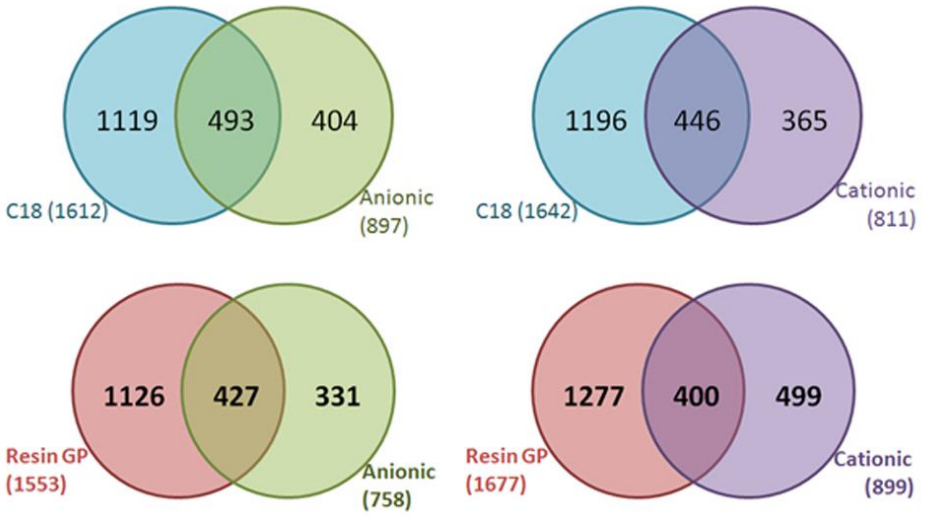


Figure 6. Venn diagram comparing molecular features (MFs) determined with a double cartridge SPE–SPE–LC–TOF/MS configuration with discrimination between the first and second cartridge for C18–anionic MM, C18–MM cationic, resin GP–anionic MM and resin GP–cationic MM.

Table 3. Molecular features detected in the SPE–SPE–LC–TOF/MS analysis with the serial configurations. The number of molecular features detected in each cartridge and the total number detected in each test are also listed. The last column represents the number of features detected with the double cartridge configuration in relation to the use of two cartridges.

Sorbent	MFs cartridge 1	MFs cartridge 2	Total MFs	Detection coverage (%)
C18–anionic MM	1612	897	2016	85.4
C18– cationic MM	1642	811	2007	92.2
Resin GP–anionic MM	1553	758	1884	79.5
Resin GP–cationic MM	1677	899	2176	99.8

Basic tentative identifications based on precursor ion are listed in Supplementary Table 1 for the different SPE–SPE couplings. This list gives an idea of the potential of this double-cartridge configuration for serum fractionation according to SPE selectivity.

4. Conclusions

The potential of SPE–LC–MS in combination with sorbents of different physico–chemical properties was assessed with a view to maximizing detection coverage in the analysis of the human serum metabolome. Sample fractionation by SPE was found to provide an effective means for enhancing the detection capabilities of a single analytical platform (LC–TOF/MS) relative to a combination of several (*e.g.* LC–MS, GC–MS, CE–MS, NMR spectroscopy). Also, the selectivity of SPE sorbents allowed serial configurations to be developed to use a combination of protocols applied in two different cartridges for the analysis of the same sample aliquot. The need for little operator intervention in the process is an additional benefit to be considered in adopting the proposed configuration for metabolomic studies involving large numbers of samples.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2012-37428). F. Priego-Capote is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks MICINN for an FPU research training scholarship (AP2009-0499). The Maimonides Biomedical Research Institute (Córdoba, Spain) is also grateful to donors of blood samples.

References

[1] B. Álvarez-Sánchez, F. Priego-Capote, M.D.L. de Castro, Trac-Trends in Analytical Chemistry 29 (2010) 111.

[2] I. Nobeli, H. Ponstingl, E.B. Krissinel, J.M. Thornton, Journal of Molecular Biology 334 (2003) 697.

[3] P.F. Suthers, M.S. Dasika, V.S. Kumar, G. Denisov, J.I. Glass, C.D. Maranas, Plos Computational Biology 5 (2009).

[4] G. Mercuro, P.P. Bassareo, M. Deidda, C. Cadeddu, L. Barberini, L. Atzori, Journal of Cardiovascular Medicine 12 (2011) 800.

[5] A.R. Fernie, R.N. Trethewey, A.J. Krotzky, L. Willmitzer, Nature Reviews Molecular Cell Biology 5 (2004) 763.

[6] M. Bedair, L.W. Sumner, Trac-Trends in Analytical Chemistry 27 (2008) 238.

[7] J. Schripsema, Phytochemical Analysis 21 (2010) 14.

[8] U. Roessner, S.G. Villas-Bôas, M.A.E. Hansen, J. Smedsgaard, J. Nielsen, in Metabolome Analysis, John Wiley & Sons, Inc., 2006, p. 15.

[9] S.M. Kang, J.C. Park, M.J. Shin, H. Lee, J. Oh, D.H. Ryu, G.S. Hwang, J.H. Chung, Clinical Biochemistry 44 (2011) 293.

[10] G.M. Liu, Y. Wang, Z.S. Wang, J.Y. Cai, X.Z. Lv, A.G. Zhou, Journal of Agricultural and Food Chemistry 59 (2011) 5572.

[11] Y. Shiomi, S. Nishiumi, M. Ooi, N. Hatano, M. Shinohara, T. Yoshie, Y. Kondo, K. Furumatsu, H. Shiomi, H. Kutsumi, T. Azuma, M. Yoshida, Inflammatory Bowel Diseases 17 (2011) 2261.

[12] A. Hirayama, K. Kami, M. Sugimoto, M. Sugawara, N. Toki, H. Onozuka, T. Kinoshita, N. Saito, A. Ochiai, M. Tomita, H. Esumi, T. Soga, Cancer Research 69 (2009) 4918.

[13] R.S. Plumb, C.L. Stumpf, J.H. Granger, J. Castro-Perez, J.N. Haselden, G.J. Dear, Rapid Communications in Mass Spectrometry 17 (2003) 2632.

[14] R. Goodacre, S. Vaidyanathan, G. Bianchi, D.B. Kell, Analyst 127 (2002) 1457.

[15] J. Smedsgaard, J. Nielsen, Journal of Experimental Botany 56 (2005) 273.

[16] O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R.N. Trethewey, L. Willmitzer, Nature Biotechnology 18 (2000) 1157.

[17] A. Ruiz-Aracama, A. Peijnenburg, J. Kleinjans, D. Jennen, J. van Delft, C. Hellfrisch, A. Lommen, Bmc Genomics 12 (2011).

[18] S.G. Villas-Boas, J. Hojer-Pedersen, M. Akesson, J. Smedsgaard, J. Nielsen, Yeast 22 (2005) 1155.

[19] R.P. Maharjan, T. Ferenci, Analytical Biochemistry 313 (2003) 145.

[20] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Analytical Chemistry 72 (2000) 3573.

[21] H. Idborg, L. Zamani, P.O. Edlund, I. Schuppe-Koistinen, S.P. Jacobsson, Journal of Chromatography B 828 (2005) 14.

[22] F. Michopoulos, G. Theodoridis, C.J. Smith, I.D. Wilson, Journal of Proteome Research 9 (2010) 3328.

[23] C.L. Winder, W.B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G.M. Stephens, R. Goodacre, Analytical Chemistry 80 (2008) 2939.

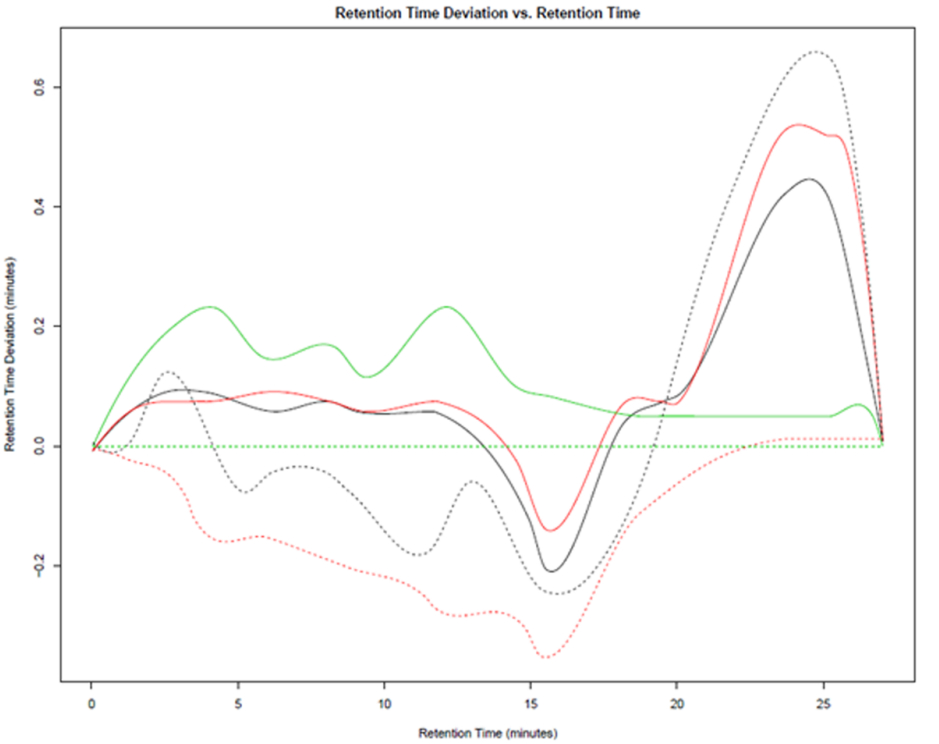
[24] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrometry Reviews 26 (2007) 51.

[25] C. Ferreiro-Vera, J.M. Mata-Granados, F. Priego-Capote, M.D. Luque de Castro, Journal of Chromatography A 1218 (2011) 2848.

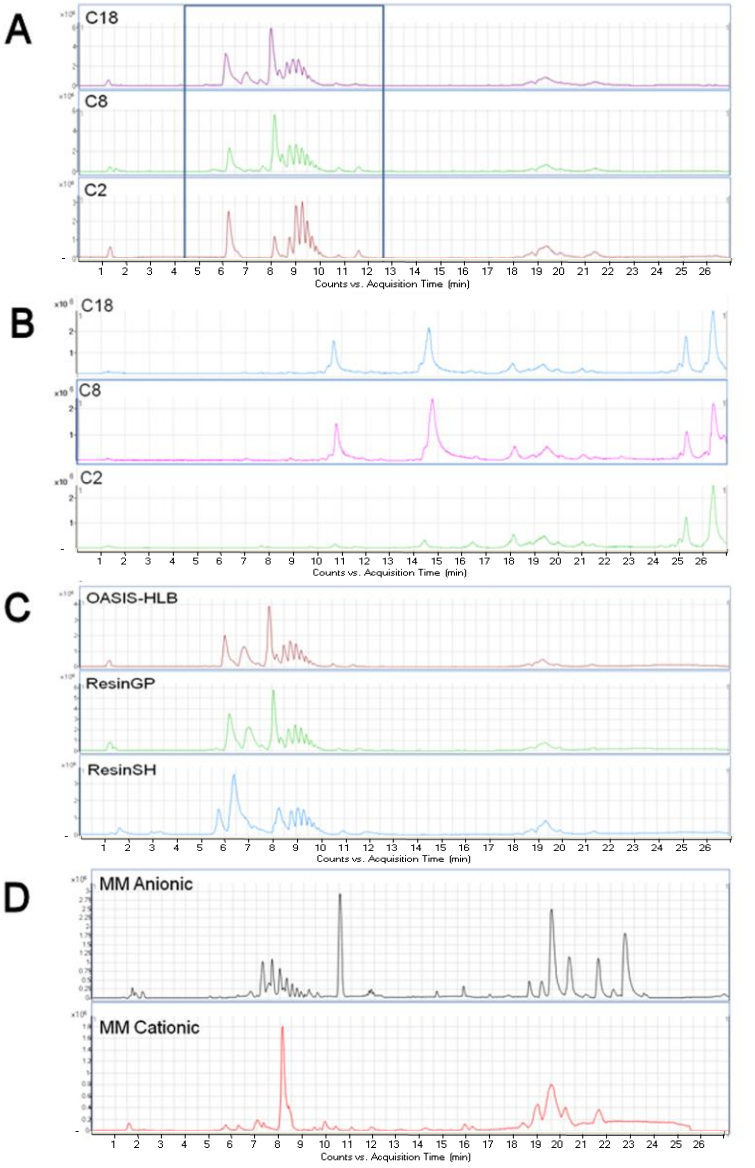
[26] C. Ferreiro-Vera, J.M. Mata-Granados, J.M. Quesada Gómez, M.D. Luque de Castro, Talanta 85 (2011) 1842.

[27] N. Psychogios, D.D. Hau, J. Peng, A.C. 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, D.S. Wishart, PLoS ONE 6 (2011) e16957.

Supplementary material



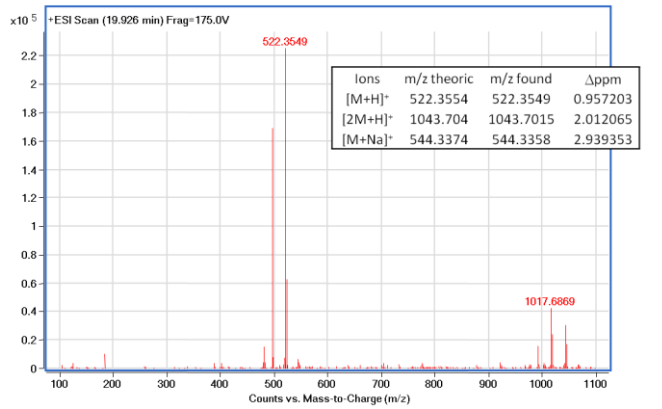
Supplementary Figure 1. Experimental variability in retention time found by analyzing three replicates of human serum by SPE–LC–TOF/MS with different sorbents.

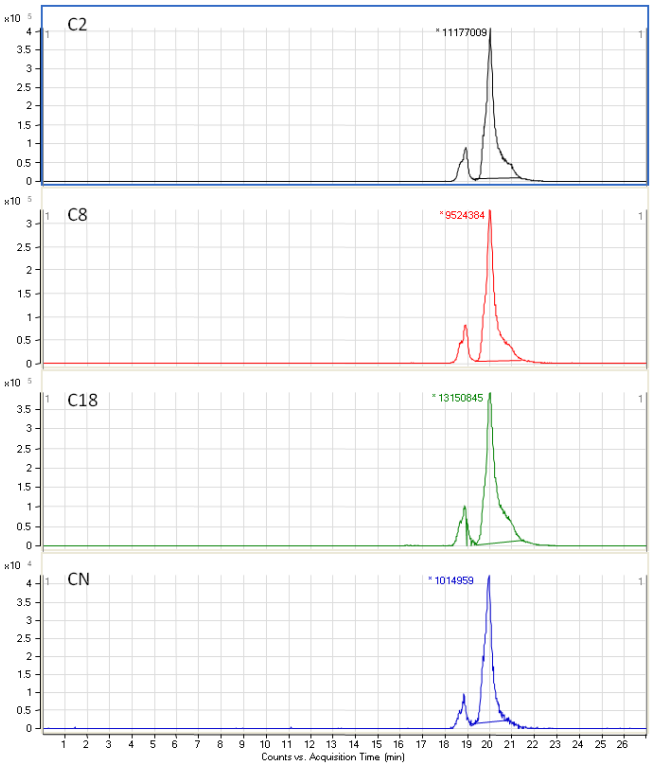


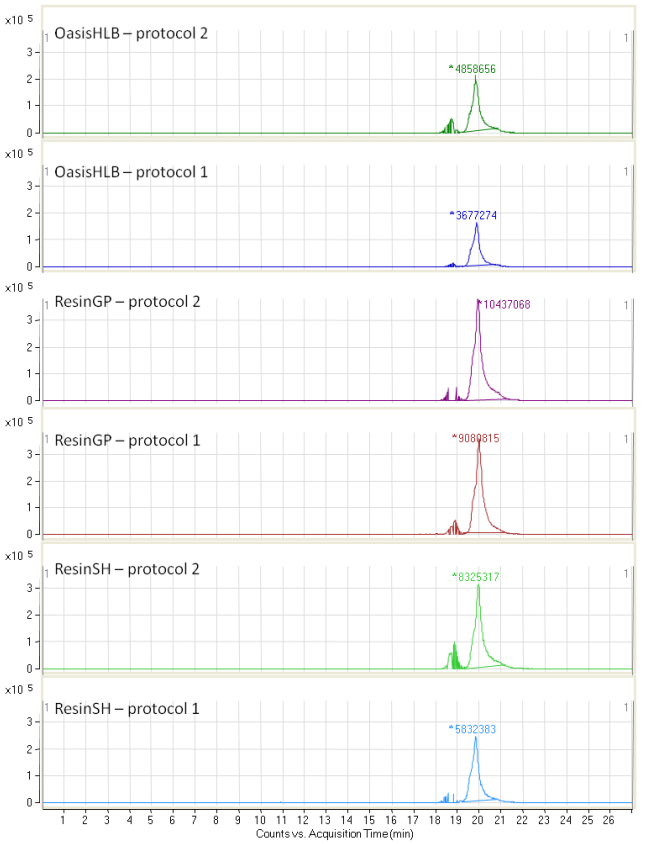
Supplementary Figure 2. Base peak chromatograms (BPCs) obtained by SPE-LC-TOF/MS analysis with the next cartridges: (A) Cartridges based on non-mixed retention mechanisms in the positive ionization mode. (B) Cartridges based on non-mixed retention mechanisms (C18, C8 and C2) in the negative ionization mode. (C) Cartridges based on mixed retention mechanisms in the positive ionization mode. (D) The two ionic sorbents in the positive ionization mode.

Lyso PC (18:1)

$C_{26}H_{52}NO_7P$ Mw=521.348145 t_R=19.9



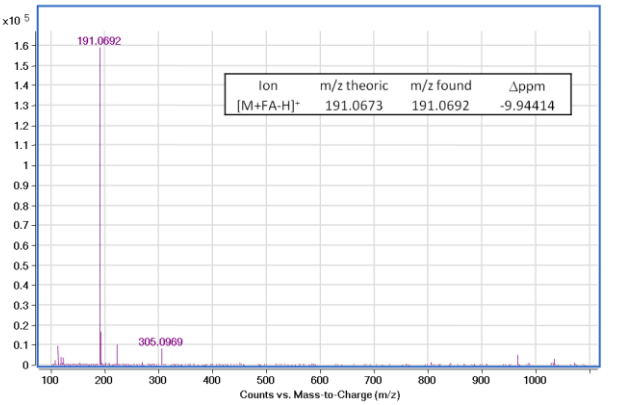


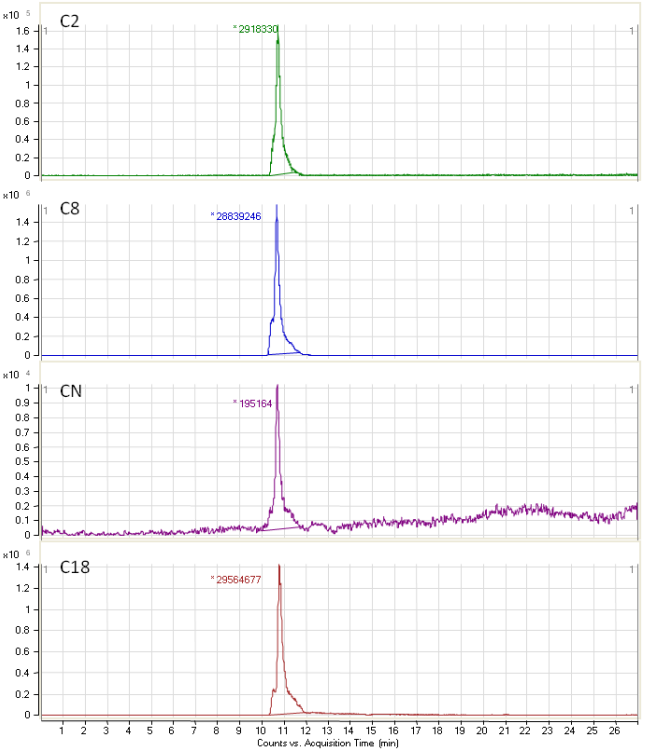


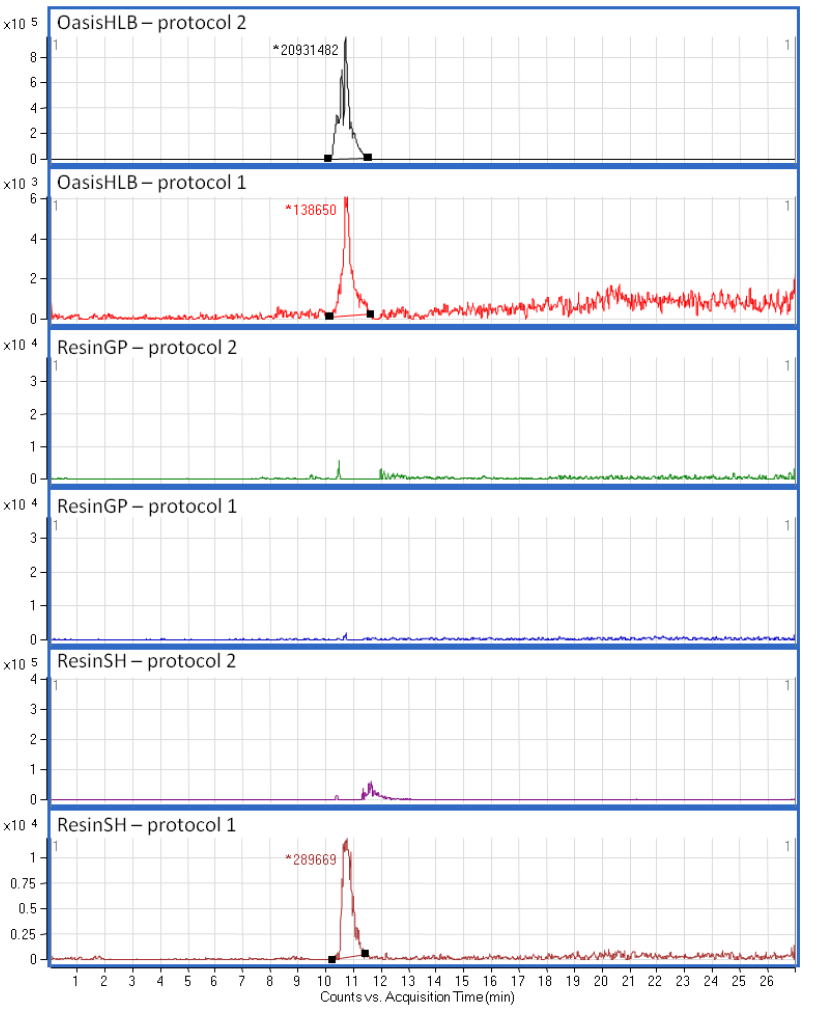
Supplementary Figure 3. A) Extracted ion chromatograms (EICs) obtained by SPE–LC–TOF/MS with different sorbents for lysophosphatidylcholine C18:1 in human serum.

Glutamine

$C_5H_{10}N_2O_3$ Mw=146.069138 t_R=10.8



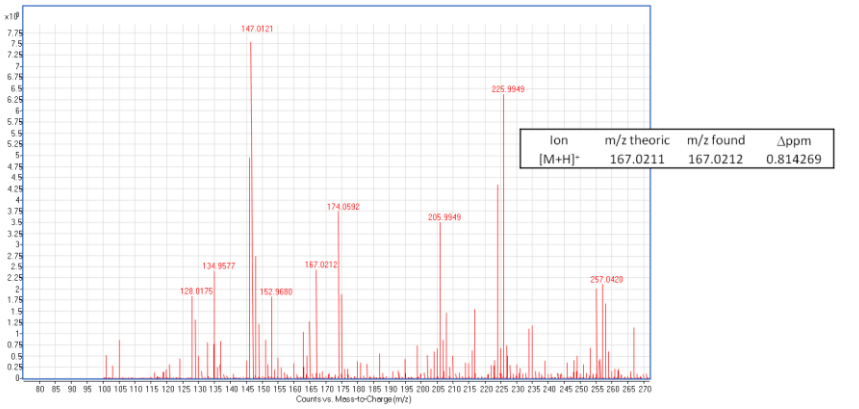


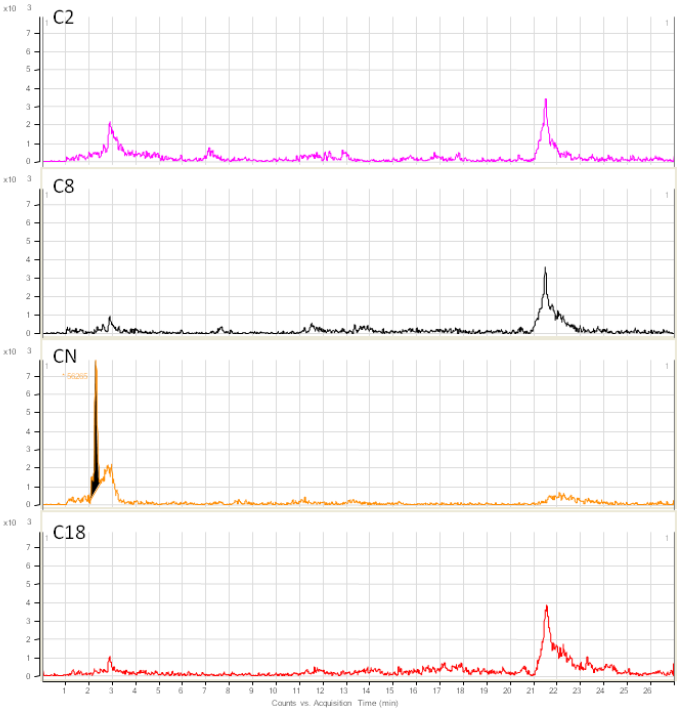


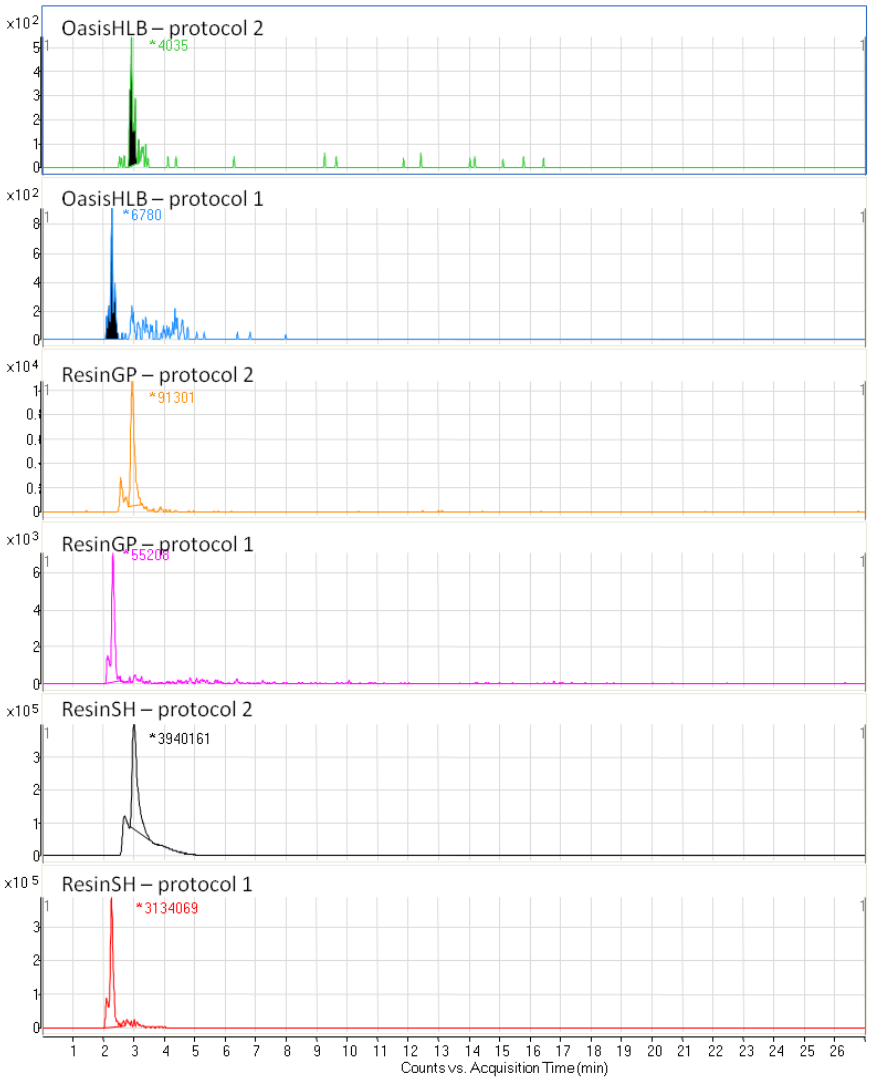
Supplementary Figure 3. C) Extracted ion chromatograms (EICs) obtained by SPE–LC–TOF/MS with different sorbents for glutamine in human serum.

Uric acid

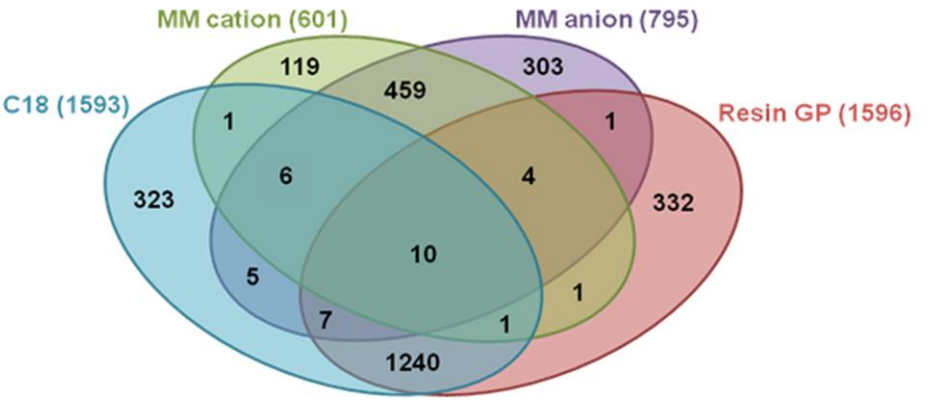
$C_5H_4N_4O_3$ Mw=168.02834 t_R=2.2



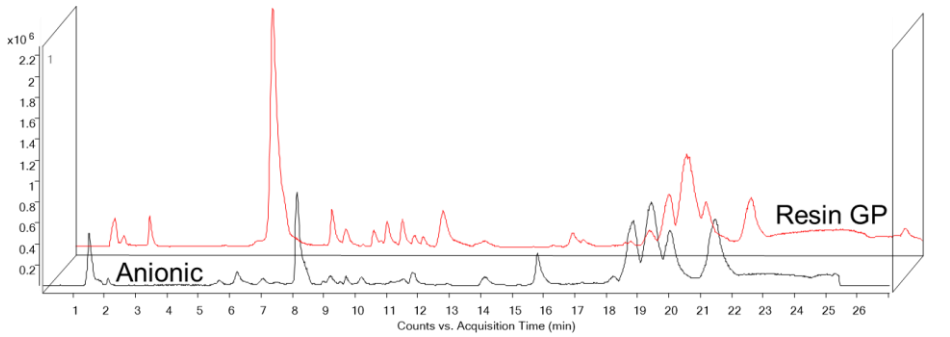




Supplementary Figure 3. D) Extracted ion chromatograms (EICs) obtained by SPE–LC–TOF/MS with different sorbents for uric acid in human serum.



Supplementary Figure 4. Venn diagram comparing molecular features (MFs) determined with a single cartridge SPE–LC–TOF/MS configuration for resin GP, C18, anionic MM and cationic MM sorbents.

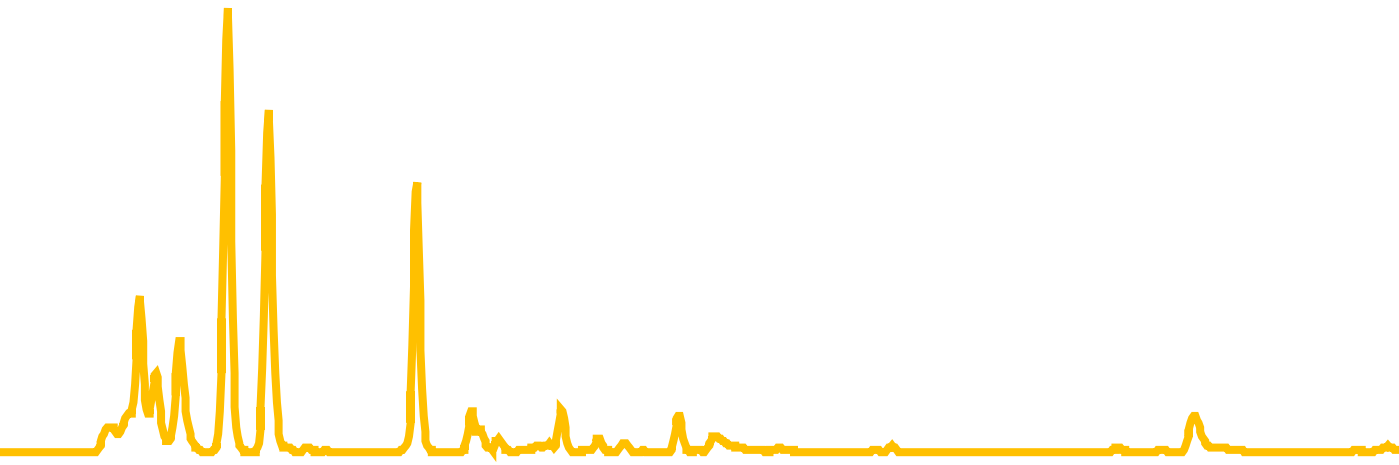


Supplementary Figure 5. Base peak chromatograms (BPCs) obtained by SPE–SPE–LC–TOF/MS analysis with resin GP–cationic MM cartridges.

Supplementary Table 1 is included in the cd-rom: "Chapter 2 – Supplementary Table 1.xls".

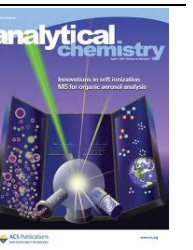
Chapter 3:

Enhanced detection and identification in metabolomics by use of LC–MS/MS untargeted analysis in combination with gas phase fractionation



Sent to

Analytical Chemistry



Enhanced detection and identification in metabolomics by use of LC–MS/MS untargeted analysis in combination with gas phase fractionation

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b]*, *M. D. Luque de Castro*[a,b]

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

Enhanced detection and identification in metabolomics by use of LC–MS/MS untargeted analysis in combination with gas phase fractionation

M. Calderón-Santiago, F. Priego-Capote, M. D. Luque de Castro*

Abstract

Liquid chromatography coupled to tandem mass spectrometry is one of the most widely used analytical platforms for profiling analysis in metabolomics. One weakness of untargeted metabolomic analysis, however, is the difficulty of identifying metabolites. In fact, the process typically involves mass-based searching of LC–MS and LC–MS/MS data, and requires using MS/MS data for unequivocal identification. Current strategies use LC–MS analysis in the scan mode prior to acquiring MS/MS information about targeted metabolites or the "auto MS/MS" mode to fragment high-abundance compounds mainly —which requires additional injections to obtain MS/MS data for minor sample components. Because an additional procedure is needed to enhance the fraction of metabolites with MS/MS data, in this work, the effectiveness of acquiring different MS/MS parameters across an analytical batch or repetitions of the same sample by using exclusion or inclusion criteria to select precursor ions is assessed. The procedure, known as "Gas Phase Fractionation (GPF)", was used here for untargeted analysis of serum. The joint use of four methods with a different mass range for selection of precursor ions each provided useful MS/MS information for at least 80% of all molecular entities detected in the MS scan replicates. By contrast, the conventional "auto MS/MS" mode of data acquisition provided MS/MS data for only 48–57% of entities and was therefore less effective toward identifying metabolites. The additional use of GPF improved the detection and identification of metabolite families such as phospholipids, amino acids, bile acids, carnitines, and fatty acids and their derivatives.

1. Introduction

The main objective of metabolomics is to identify and/or quantify small molecules or metabolites (<1500 Da) present in biofluids, cells or organisms.[1] Metabolomics uses three different strategies mainly (viz., fingerprinting, targeted and untargeted or profiling analysis),[2] but especially the latter two. Since the fingerprinting strategy is aimed primarily at the analysis of specific families or groups of metabolites, the underlying analytical methodology can be easily adapted to obtain the sensitivity and specificity levels required for each individual application. On the other hand, untargeted (profiling) analysis is aimed at identifying the greatest possible number of metabolites with a single, high-throughput analytical protocol or a combination of complementary protocols if needed. Although no single analytical tool encompassing the whole metabolome is currently available,[3] recent advances in liquid chromatography coupled to mass spectrometry (LC–MS) have turned this analytical tool into an effective platform for the unbiased detection of a large number of metabolites from biological samples.[4–8] In fact, LC–MS/MS has considerably facilitated metabolite identification and quantitation by improving MS sensitivity and precision.[9]

By virtue of its role as a primary carrier of metabolites in the human body, serum/plasma is the most common target biofluid of clinical and nutritional metabolomics analyses. Serum/plasma provides a liquid highway for all molecules secreted, excreted or discarded by different tissues in response to different physiological needs or stresses. However, the high complexity of serum hinders obtaining a complete metabolite profile in a single analysis. In fact, characterizing the serum metabolome has required using a combination of five different instrumental platforms.[3]

At present, metabolite identification by untargeted analysis is accomplished mainly via mass-based searches of LC–MS and LC–MS/MS data, the latter being essential for unequivocal identification. In fact, successful identification of metabolites requires managing comprehensive MS/MS spectral

libraries and MS/MS information for each potential molecular entity detected in a metabolomic analysis.

The MS/MS information needed can be obtained by manual selection of precursors based on LC–MS data or by use of the "auto MS/MS" acquisition mode, which is much less time-consuming.[10] This mode combines MS scan cycles with programmed MS/MS scans of selected precursor ions depending on their relative abundance in the MS scan, thereby allowing both MS and MS/MS data to be acquired in a single analysis. However, not all metabolites detected by MS scanning will be fragmented and give an MS/MS spectral signal, so fragmentation of the most abundant metabolites —or those with high ionization capabilities— is to be preferred over that of less concentrated metabolites.

Applying different MS/MS acquisition parameters based on preset exclusion or inclusion criteria for selection of precursor ions to an analytical batch or to replicates of a sample can increase the fraction of metabolites providing useful MS/MS data. This procedure, termed "Gas Phase Fractionation" (GPF), which has been validated for proteomics applications,[11-13] has substantially enhanced the identification of proteins. GPF has been introduced in metabolomics by developing an algorithm that provides the set of methods needed to obtain MS/MS spectra for a ranked list of interesting precursor features compiled by automated selection of quasi-molecular ions after LC–MS profiling.[14] The underlying methodology requires injection of the sample in the MS and MS/MS modes.

The aim of this work was to assess the usefulness of GPF to increase the level of information obtained by LC–MS/MS analysis of serum. For this purpose, a combination of methods based on different inclusion criteria for selection of precursor ions to undergo MS/MS fragmentation was programmed in order to assess the influence of GPF on metabolomic coverage. Since accurate metabolomic profiling requires using at least three replicates per sample,[15] using GPF in combination with appropriate replicates is an effective approach to enhancing identification capabilities without the need for further sample injections.

2. Experimental section

2.1. Chemicals

Chromatographic mobile phase B was prepared in LC–MS grade acetonitrile from Sigma–Aldrich (Madrid, Spain). MS grade formic acid from Scharlab (Barcelona, Spain) was used as ionization agent in LC–QTOF analyses, and deionized water (18 MΩ·cm) supplied by a Millipore Milli-Q water purification system from Millipore (Bedford, MA, USA) was used to prepare the chromatographic aqueous phase. All samples were prepared in LC grade methanol from Scharlab.

2.2. Instruments and apparatus

A Sorvall Legend Micro 21R centrifuge from Thermo Scientific (Waltham, MA, USA) was used to centrifuge serum samples after protein precipitation. An Agilent 1200 Series LC system consisting of a binary pump, a vacuum degasser, an autosampler and a thermostated column compartment, and coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with a dual electrospray (ESI) source (Santa Clara, CA, USA), was used to obtain MS spectra. Chromatographic eluates were monitored by tandem mass spectrometry in the high resolution mode.

2.3. Blood collection and serum isolation

All steps from blood sampling to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004) and supervised by specialist staff of the Maimonides Biomedical Research Institute (Cordoba, Spain).

Venous blood was collected from ten healthy individuals into evacuated sterile serum tubes containing no additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated at room temperature for 30 min to facilitate coagulation. Then, the tubes were centrifuged at 4 °C at 2000 g for 15 min to isolate the serum fraction, which was pooled (200 μL per participant) prior to placing in plastic tubes and storage at –80 °C until analysis.

2.4. Sample preparation

Aliquots of 100 μL of serum pool were immersed in an ice bath and deproteinized with 200 μL of methanol.[16] The mixture was shaken for 1 min and the resulting precipitate removed after centrifugation at 4 °C at 13 800 g for 5 min. Then, the upper phase was collected in a vial and the vial placed on the chromatograph autosampler for subsequent analysis.

2.5. LC–QTOF MS/MS analysis

Chromatographic separation was performed by using a Mediterranean C18 reversed phase analytical column (50 mm × 0.46 mm i.d., 3 μm particle size) from Teknokroma (Barcelona, Spain) which was thermostated at 25 °C. The mobile phases were water (phase A) and ACN (phase B), both containing 0.1% formic acid as ionization agent. The LC pump was programmed to use a flow rate of 0.8 mL/min and the following elution gradient: 3% phase B as initial mobile phase (2 min), ramp to 100% phase B (12 min) and holding (3 min). A post-time of 3 min was used to regain the initial conditions for the next analysis. The injected volume was 5 μL and the injector needle was washed 10 times with 80% methanol between injections. Also, the needle seat back was flushed with 80% methanol at 4 mL/min 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 ±4.0 kV, fragmentor voltage 175 V, N_2 pressure in the nebulizer 40 psi; N_2 flow rate and temperature as drying gas 10 L/min and 350 °C, respectively. The instrument was calibrated and tuned as recommended by the manufacturer. MS/MS data were acquired in both polarities, using the centroid mode at a rate of 2.5 spectra/s in the extended dynamic range mode (2 GHz). Accurate mass spectra over the *m/z* range 100–1000 were acquired in the MS scan mode, and MS/MS was performed with automated selection of 2 precursor ions per cycle and an exclusion window of 0.25 min after 2 consecutive selections of the same precursor. The collision energy was set at 20 V for the whole run.

Five experimental protocols were designed by restricting the inclusion range for selection of precursor ions in the MS/MS acquisition range as shown in Table 1. A total of 12 replicates of deproteinized serum were analyzed in each experimental protocol. The first experimental protocol used a conventional "auto MS/MS" method and the inclusion *m/z* range from 100 to 1000. The second experimental protocol was based on a combination of two MS/MS methods differing in the inclusion range used to select precursor ions. Thus, 6 replicates were analyzed over the *m/z* range 100–550 and another 6 over the range 550–1000. The third and fourth experimental protocols used a combination of 3 and 4 methods, respectively, and 3 and 4 different inclusion ranges (viz., *m/z* 100–400, 400–700 and 700–1000 with the third protocol; and *m/z* 100–325, 325–550, 550–775 and 775–1000 with for the fourth), respectively. Finally, the fifth protocol involved a combination of six methods with different inclusion ranges. A total of 60 analyses were thus conducted on the set of replicates.

The QTOF spectrometer provided a typical resolution of 15 000 FWHM (Full Width at Half Maximum) at *m/z* 118.086255 and 30 000 FWHM at *m/z* 922.009798. In order to assure the required mass accuracy for recorded ions, the spectrometer was subjected to continuous internal calibration during analyses by using the signals at *m/z* 121.0509 (protonated purine) and 922.0098 [protonated hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazine or HP-921] in the positive ion mode; and those at *m/z* 119.0362 (proton abstracted purine) and 966.0007 (formate adduct of HP-921) in the negative ion mode.

2.6. Data processing and statistical analysis

The MassHunter Workstation software package (B.05.00 Qualitative Analysis and B.06.00 Profinder, Agilent Technologies, Santa Clara, CA, USA) 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. This algorithm initially deconvolutes chromatograms and aligns features across the selected sample files in terms of mass and retention time; then, it uses the mass

and retention time of each feature for recursive targeted feature 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 and a maximum charge state of 2. In addition, the isotopic distribution for a 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 in the positive (+H, +Na) and negative ionization mode (–H, +HCOO) was also used to identify features of the same molecule. Features were aligned by using a tolerance window of 0.30 min and a mass accuracy of 10 ppm for retention times and m/z values across all data files, respectively. The minimum absolute height required for feature extraction was set at 3000 counts, which was also used for 100% of samples in the recursive step.

Once all features were extracted and aligned, the software MassHunter Qualitative was used for the targeted extraction of MS/MS information from all molecular features in the whole set of analyses. The most salient entities were identified by using MS and MS/MS information, and searching the METLIN MS and MS/MS database (http://metlin.scripps.edu) and the Human Metabolome Database (HMDB, version 3.5).

3. Results and discussion

3.1. Common LC–MS serum elution profiles

As noted earlier, only the molecular features detected in the 60 replicates of serum analyses were considered. Following molecular feature extraction and alignment, two data sets consisting of 139 and 158 molecular entities in the negative and positive ionization mode, respectively, were obtained. Serum is a complex biofluid containing water-soluble metabolites and a significant lipid fraction as shown by the LC–MS/MS elution profiles obtained in the positive and negative ionization mode (see Figure 1). As can be seen, the elution profiles revealed the presence of molecular features with variable m/z values and

retention times. MS and MS/MS information allowed most of the metabolites detected in each ionization mode to be identified. Those identified were grouped into chemical families, some of which were detected in both ionization modes.

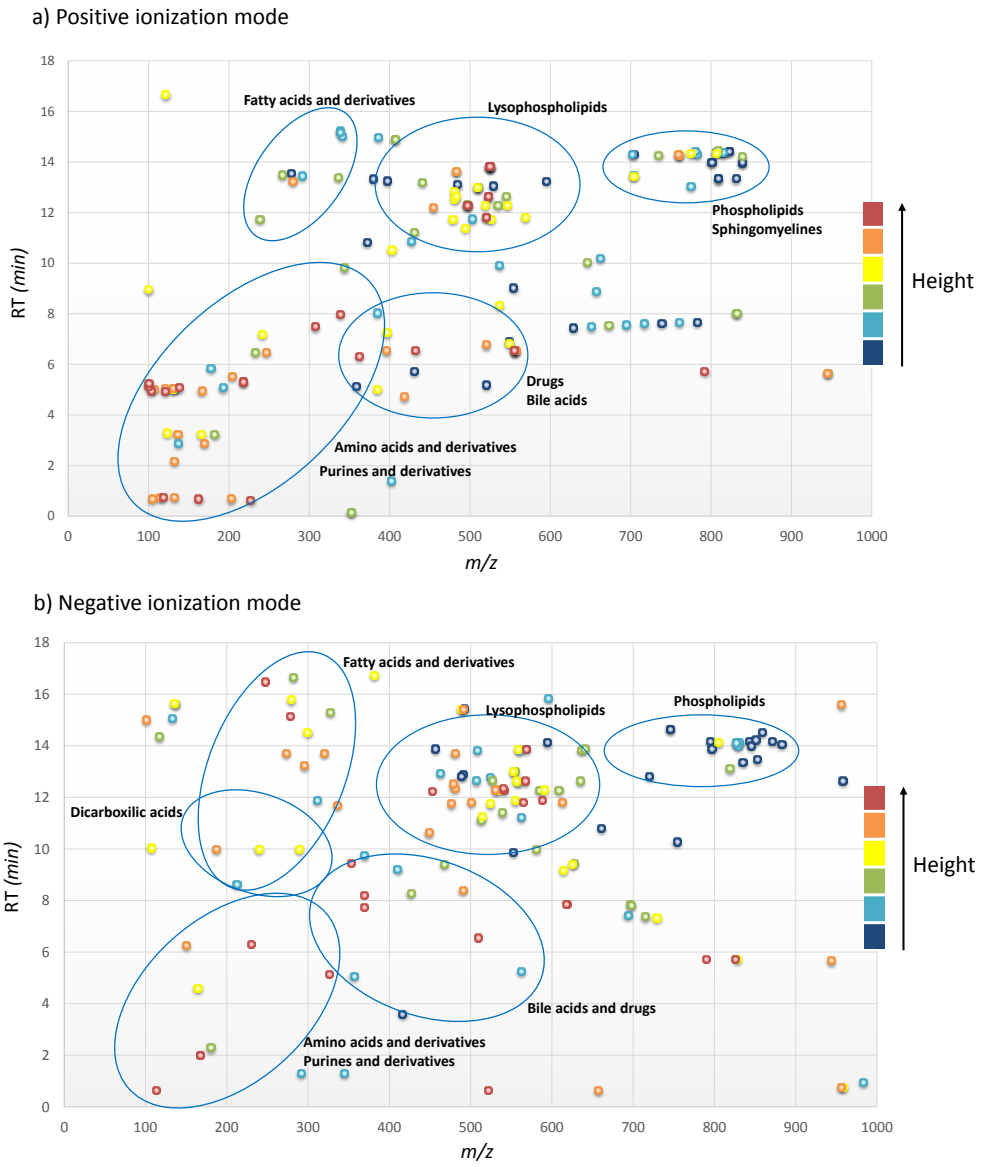


Figure 1. Chromatographic profiles obtained by LC–MS/MS analysis of serum in the positive (a) and negative ionization mode (b).

Table 1. Experimental protocols for the LC–MS/MS analysis of human serum.

Protocol	Method	MS scan (*m/z*)	Inclusion range (*m/z*)	Number of replicates
1	1	100–1000	100–1000	12
2	2.1	100–1000	100–550	6
	2.2		550–1000	6
3	3.1	100–1000	100–400	4
	3.2		400–700	4
	3.3		700–1000	4
4	4.1	100–1000	100–325	3
	4.2		325–550	3
	4.3		550–775	3
	4.4		775–1000	3
5	6.1	100–1000	100–250	2
	6.2		250–400	2
	6.3		400–550	2
	6.4		550–700	2
	6.5		700–850	2
	6.6		850–1000	2

Worth special note among the metabolites identified in the negative ionization mode were nonpolar compounds such as phospholipids, which contain ionizing functional groups or tend to form adducts with formic acid. Compounds containing carboxyl groups such as fatty acids, bile acids and dicarboxylic acids were also prevalent in the elution profile for the negative ionization mode. On the other hand, the elution profile for the positive ionization mode was consistent with typical polar metabolites such as amino acids and purines, and also with their derivatives —the amino group is easily ionized in acid media. Bile acids and some phospholipids were also unequivocally identified in the elution profile for the positive ionization mode by virtue of most having some amino group. Interestingly, single-chain phospholipids (particularly phosphatidylcholines and phosphatidylethanolamines) were successfully identified in the positive ionization mode thanks to their characteristic fragmentation pattern. On the

other hand, double-chain phospholipids required using MS/MS spectra in the negative ionization mode to confirm their chain length. Table S-1 lists most of the metabolites identified in serum as classified into chemical families.

Both elution profiles contained a large number of metabolites coeluting over the range 10–16 min, which was confirmed by the times in the positive ionization mode: 0–8 min. Thus, a significant fraction of metabolites had similar retention times but differed in signal intensity by effect of their different concentrations and/or ionization capabilities. In fact, most of the tentative metabolites detected from low signals coeluted with metabolites giving higher signals. One can therefore assume that the precursor selection process used in Protocol 1 ("auto MS/MS" mode, Table 1) favored the detection of abundant metabolites or compounds with a high ionization power over that of metabolites with a low concentration or ionization capability, which were not fragmented.

3.2. Comparison of MS/MS detection coverage in terms of the acquisition method

The purpose of this test was to compare MS/MS detection coverage by using a single acquisition method in combination with GPF methods involving different inclusion ranges for selection of precursor ions. The reference method used for this purpose was that based on a single MS/MS inclusion range (m/z 100–1000). Figure 2 shows the distribution of tentative metabolites obtained by using this method with MS/MS information in the positive and negative ionization mode as a function of m/z.

Although a total of 158 and 139 molecular entities were detected and aligned in the positive and negative ionization mode, respectively, only 90 and 67 tentative metabolites, respectively, present in all replicates were fragmented by MS/MS with the reference method. Therefore, only 57% of all entities were detected in the positive mode and 48% in the negative mode. As can be seen from Figure 2, a high proportion of molecular features was detected over the *m/z* range 400–550 —which corresponds to lysophospholipids mainly— in the positive mode and over the range 400–700 —potentially associated with phospholipids,

which possess easily ionized functional groups and are highly concentrated in serum— in the negative mode. Figure 3 shows the number of tentative metabolites fragmented by MS/MS with each method used in the five protocols according to the inclusion ranges for precursor selection. As can be seen, the results were similar in both ionization modes; in fact, the number of tentative metabolites detected by MS/MS increased with that of methods used by the GPF protocol, but was slightly smaller with the protocol using a combination of 6 MS/MS inclusion ranges.

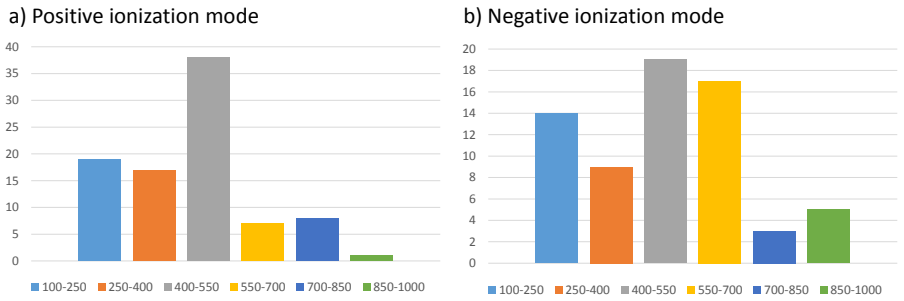


Figure 2. Distribution of tentative metabolites with MS/MS information in serum in the positive (**a**) and negative ionization mode (**b**) as a function of *m/z*.

Thus, the proportion of tentative metabolites detected from MS/MS information by using the 2-method GPF protocol increased to 73 and 63% of the total number of entities identified after alignment. The maximum coverage was obtained with the 4-method GPF protocol, which allowed MS/MS information for about 80% of all tentative metabolites included in the original data set to be extracted. A comparison of the protocols using a combination of 2 and 4 methods revealed that the latter allowed a greater number of tentative metabolites to be identified over the *m/z* range 100–325 in the positive ionization mode (particularly amino acids, purines and their derivatives, which are very easily ionized in this mode). By contrast, the greatest increase in MS/MS acquisitions with the negative ionization mode was obtained in the *m/z* range 550–1000, which was assigned to phospholipids. Based on these results, GPF protocols for

metabolomic analysis provide an effective tool for expanding metabolite coverage (particularly with coeluting compounds).

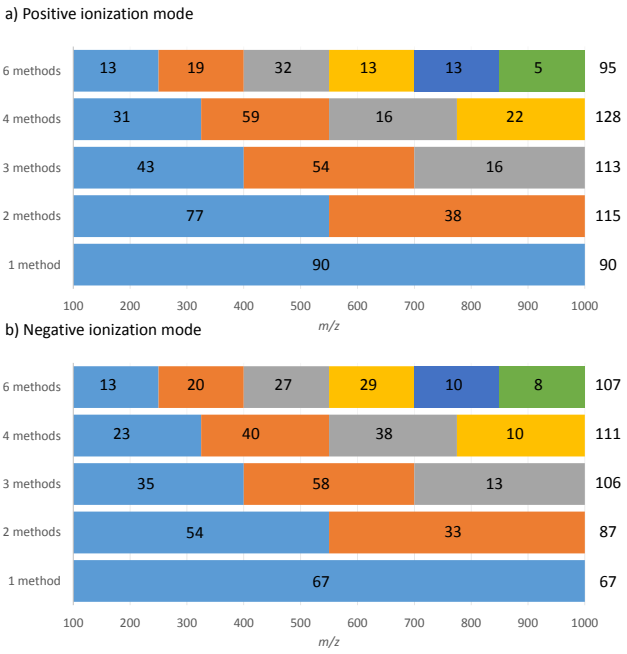


Figure 3. Number of tentative metabolites fragmented with different LC–MS/MS protocols using a single method or GPF protocols based on a combination of 2–6 methods. (**a**) Positive ionization mode. (**b**) Negative ionization mode.

3.3. Identification of metabolites in the positive ionization mode

As noted earlier, some metabolite families (particularly amino acids, acylcarnitines, purines, lysophospholipids and sphingomyelins) are preferentially detected in the positive ionization mode with LC–MS/MS. As stated in the previous section, metabolites with m/z < 550 were the most benefited by the use of precursor ion restricted fragmentation in GPF. Figure 4 shows the specific metabolites identified in each compound family as a function of the GPF acquisition method used. Several amino acids were easily identified in all replicates with the five protocols, namely: betaine, carnitine, phenylalanine, phenylglycine and tryptophan. Creatine was detected with all other protocols and

the amino acids (iso)leucine and tyrosine were only detected with combinations of 2, 3 and 4 GPF methods. As can be seen from Figure S-1, the extracted ion chromatograms (EICs) for these amino acids overlapped over a narrow elution interval [0.8–1.1 min for (iso)leucine/and 1.8–2.2 min for tyrosine]; as a result, they could only be detected with some GPF methods. A similar behavior was observed in purine and its derivatives. Thus, purine was clearly identified with both GPF and the conventional acquisition method, but two derivative metabolites giving higher peak areas (viz., uric acid and hypoxanthine) were only detected with the GPF protocol using 2, 3 and 4 MS/MS acquisition methods in combination (see EICs in Figure S-2).

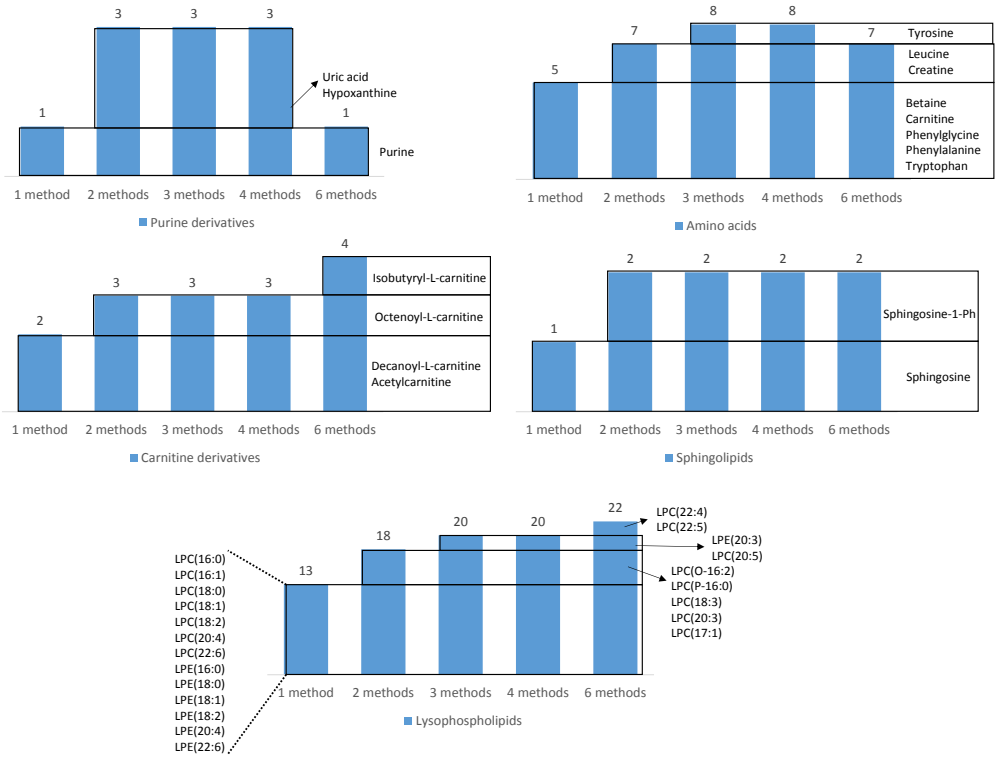


Figure 4. Number of compounds with MS/MS information for the most important families identified in positive ionization mode as a function of the tested protocol.

Lysophospholipids constitute one other family of compounds that is easily identified in the positive ionization mode. Lysophosphatidylcholines (LPCs) and lysophosphatidylethanolamines (LPEs), which are the most common lysophospholipids, exhibit a characteristic fragmentation pattern in the positive ionization mode.[17] This family was detected over the mass range *m/z* 400–700 in the elution window from 11 to 14 min; a total of 13 lysophospholipids including the most common or abundant LPCs and LPEs were thus identified by using the conventional MS/MS acquisition method. Using 2 GPF methods in combination increased the number of lysophospholipids detected to 18, and using 3 or 4 methods expanded coverage to 20 compounds of this family. Finally, using 6 GPF methods allowed the identification of 22 LPCs and LPEs including less common LPCs such as LPC(O-16:2) and the plasmalogen LPC(P-16:0), which form by condensation with an alkyl alcohol (C16:2) and by etherification with glycerol, respectively.

One other example of the benefits of GPF methodology was that of sphingolipids. Thus, sphingosine was clearly identified in all replicates, both with GPF and with the conventional method. However, sphingosine-1-phosphate was only identified by GPF. Carnitine, present in free and esterified forms in human serum, is essentially involved in the metabolism of free fatty acids. As stated above, carnitine in free form was successfully detected irrespective of the MS/MS acquisition method used. However, only acetylcarnitine and decanoylcarnitine among acylcarnitines were identified with the conventional acquisition method — by contrast, GPF enabled identification of two additional acylcarnitines. Thus, octanoylcarnitine was identified with all GPF protocols, and isobutyrylcarnitine with 6 MS/MS acquisition ranges only (see Figure S-3). In this particular case, the narrower were the mass ranges used for selection of precursor ions, the better was identification of low-molecular weight carnitine analogues.

3.4. Identification of metabolites in the negative ionization mode

The main effect of using GPF in the negative ionization mode was increasing the number of metabolites identified at m/z values above 550. This

spectral region corresponds mainly to phospholipids (particularly phosphatidyl-cholines and phosphatidylethanolamines, which are the two most concentrated subclasses of glycerophospholipids in human serum). Figure 5 compares the number of metabolites identified with different methods in the negative ionization mode. No phospholipids could be identified with the conventional MS/MS acquisition method. By contrast, GPF allowed several phosphatidylcholines to be identified (particularly by use of 6 different mass ranges to select precursor ions). Identification of fatty acids and related compounds was similarly affected by the use of GPF, which allowed 3–9 compounds in this family to be identified. The conventional MS/MS acquisition method enabled the identification of 12-HETE, 9-HODE and leukotriene B_4, which form by oxidation of essential fatty acids and contain easily ionized functional groups. Using GPF methodology allowed many other major fatty acids such as oleic, linoleic, linolenic acid, arachidonic and docosahexanoic, in addition to the minor metabolite 13-HpODE —which eluted near 9-HODE and 12-HETE— to be identified. As can be seen from the extracted ion chromatograms for fatty acids in Figure S-4, all free fatty acids identified eluted at times from 15 to 17 min, which is consistent with the fragmentation pattern observed by using GPF methodology.

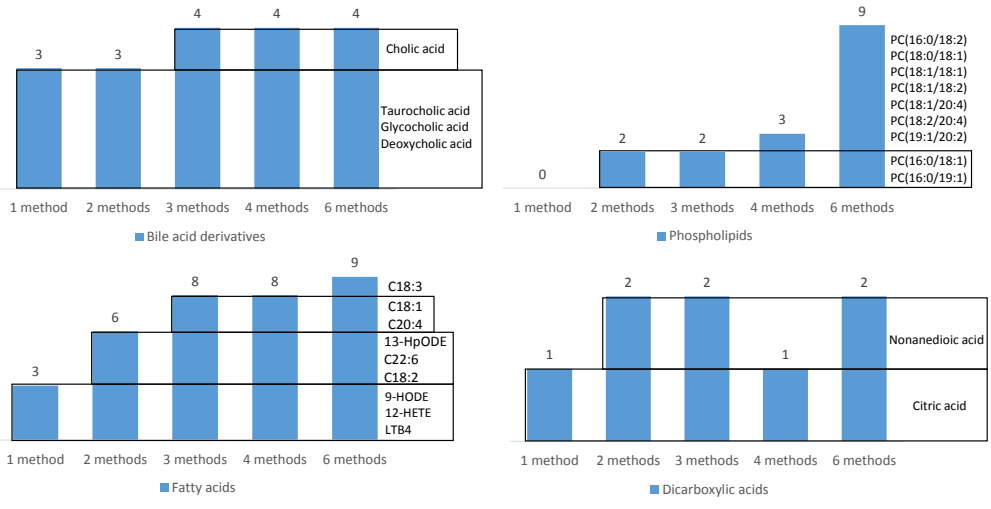


Figure 5. Number of compounds with MS/MS information for the most important families identified in negative ionization mode as a function of the tested protocol.

One other family of metabolites identified in the negative ionization mode was that of bile acids. Thus, taurocholic, glycocholic and deoxycholic acid were identified with both GPF and the conventional MS/MS acquisition method. By contrast, cholic acid could only be identified with the GPF protocols based on a combination of 3 or more mass ranges for selection of precursor ions because its peak was shorter than those for the other bile acids (see Figure S-5). The dicarboxylic acids citric and nonanedioic were also identified in human serum. Finally, citric acid was identified with all methods, but the MS/MS spectrum for nonanedioic acid could only be acquired with GPF.

4. Conclusions

Poor metabolite identification is one of the main current weaknesses of profiling analysis in metabolomics owing to the absence of a comprehensive MS/MS database and the difficulty of obtaining MS/MS information for most detected metabolites.

Using GPF methodology for serum analysis by LC–MS/MS in metabolomics has proved an effective strategy for increasing the amount of MS/MS information available about some compounds. A combined method splitting the range of precursor ion masses into four intervals was found to be best the choice among all studied here as it provided MS/MS information for at least 80% of all detected entities. By contrast, the conventional auto MS/MS mode of data acquisition provided information for only 48–57% and was therefore less effective for unequivocal identification of metabolites.

Because samples are usually injected at least in triplicate in metabolomic analyses, the proposed methodology can be easily applied to sample batches without the need for a previous LC–MS injection. A combination of three methods with different exclusion intervals for selection of precursor ions can be

used to expand MS/MS information without the need for repeated injection of samples. Also, a combination of appropriate exclusion intervals suited to the target compounds can be used to maximize the number of potential molecular features extracted from MS/MS data.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2012-37428). F. Priego-Capote is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks the MICINN for an FPU scholarship (AP2009-0499). The Maimonides Biomedical Research Institute (Córdoba, Spain) is also grateful to donors of blood samples.

References

(1) Bowen, B.P.; Northen, T.R. *J. Am. Soc. Mass Spectrom.* 2010, *21*, 1471–1476.

(2) Álvarez-Sánchez, B.; Priego-Capote, F.; Luque de Castro, M.D. *Trends Anal. Chem.* 2010, *29*, 120–127.

(3) Psychogios, N.; Hau, D.D.; Peng, J.; Guo, A.C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner,R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T.L.; Smith, S.R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J.W.; Goodfriend, T.; Wishart, D.S. *PLoS ONE* 2011, *6*, e16957.

(4) Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 2012, *883-884*, 68–75.

(5) Metz, T.O.; Zhang, Q.; Page, J.S.; Shen, Y.; Callister, S.J.; Jacobs, J.M.; Smith, R.D. *Biomark. Med.* 2007, *1*, 159–185.

(6) Roux, A.; Lison, D.; Junot, C.; Heilier, J.F. *Clin. Biochem.* 2011, *44*, 119–135.

(7) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. *Analyst* 2012, *137*, 293–300.

(8) Zhou, B.; Xiao, J.F.; Tuli, L.; Ressom, H.W. *Mol. Biosyst.* 2012, *8*, 470–481.

(9) Xiao, J.F.; Zhou, B.; Ressom, H.W. *Trends Anal. Chem.* 2012, *32*, 1–14.

(10) Dunn, W.B.; Erban, A.; Weber, R.J.M.; Creek, D.J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M.R. *Metabolomics* 2013, *9*, S44–S66.

(11) Kennedy, J.; Yi, E.C. *Methods Mol. Biol.* 2008, *432*, 217–228.

(12) Blonder, J.; Rodriguez-Galan, M.C.; Lucas, D.A.; Young, H.A.; Isaaq, H.J.; Veenstra, T.D.; Conrads, T.P. *BBA – Proteins Proteom.* 2004*, 1698*, 87–95.

(13) Scherl, A.; Shaffer, S.A.; Taylor, G.K.; Kulasekara, H.D.; Miller, S.I.; Goodlett, D.R. *Anal. Chem.* 2008, *80*, 1182–1191.

(14) Neumann, S.; Thum, A.; Böttcher, C. *Metabolomics* 2013, *9*, S84–S91.

(15) Korman, A.; Oh, A.; Raskind, A.; Banks, D. *Methods Mol. Biol.* 2012, *856*, 381–413.

(16) Bruce, S.J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P.A. *Anal. Chem.* 2009, *81*, 3285–3296.

(17) Pulfer, M.; Murphy, R.C. *Mass Spectrom. Rev.* 2003, *22*, 332–364.

Supplementary material

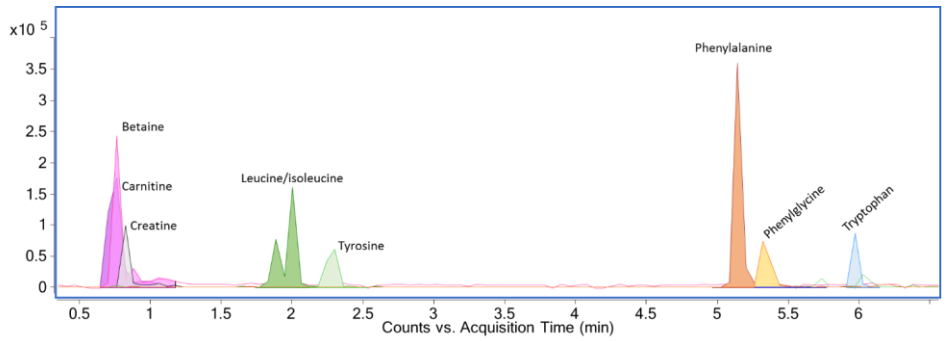


Figure S-1. Extracted ion chromatograms (EICs), in positive ionization mode, of the identified amino acids: betaine (118.0867 *m/z*), carnitine (162.1124 *m/z*), creatine (132.0769 *m/z*), leucine/isoleucine (132.1016 *m/z*), tyrosine (182.0812 *m/z*), phenylalanine (166.0856 *m/z*), phenylglycine (152.0706 *m/z*), and tryptophan (205.0964 *m/z*).

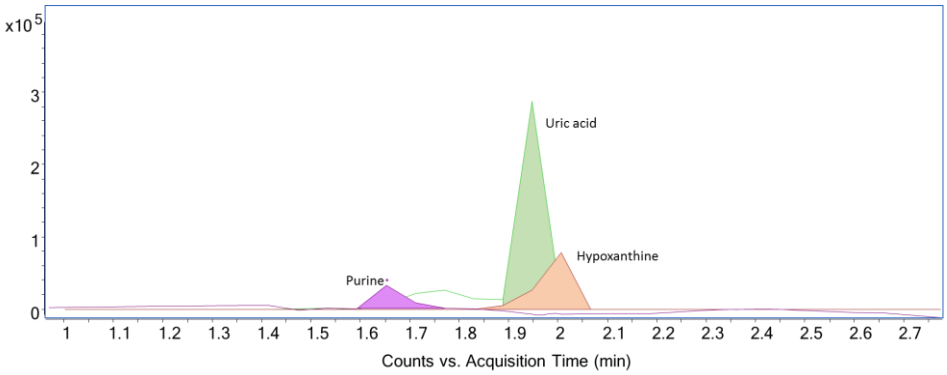


Figure S-2. Extracted ion chromatograms (EICs), in positive ionization mode, of purine (121.0509 *m/z*) and its identified derivatives: uric acid (169.0354 *m/z*) and hypoxanthine (137.0457 *m/z*).

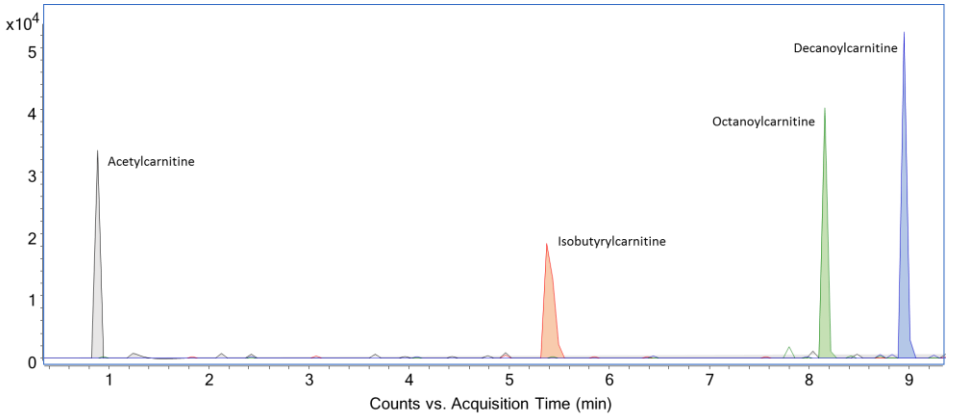


Figure S-3. Extracted ion chromatograms (EICs), in positive ionization mode, of the identified acylcarnitines: acetylcarnitine (204.1234 *m/z*), isobutyrylcarnitine (232.1536 *m/z*), octanoylcarnitine (288.2164 *m/z*) and decanoylcarnitine (316.2486 *m/z*).

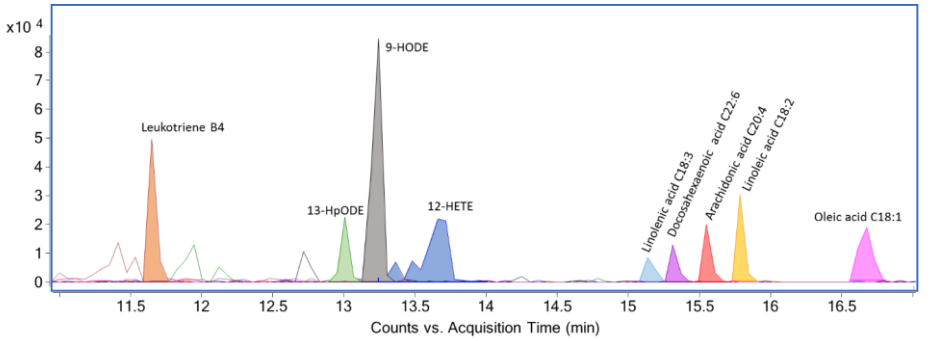


Figure S-4. Extracted ion chromatograms (EICs) of fatty acids and relative metabolites identified in negative ionization mode: leukotriene B4 (335.2244 *m/z*), 13-HpODE (311.2246 *m/z*), 9-HODE (295.2305 *m/z*), 12-HETE (319.2295 *m/z*), linoleic acid (279.2354 *m/z*), docosahexaenoic acid (327.2354 *m/z*), arachidonic acid (303.2336 *m/z*), linolenic acid (277.2173 *m/z*), and oleic acid (281.2517 *m/z*).

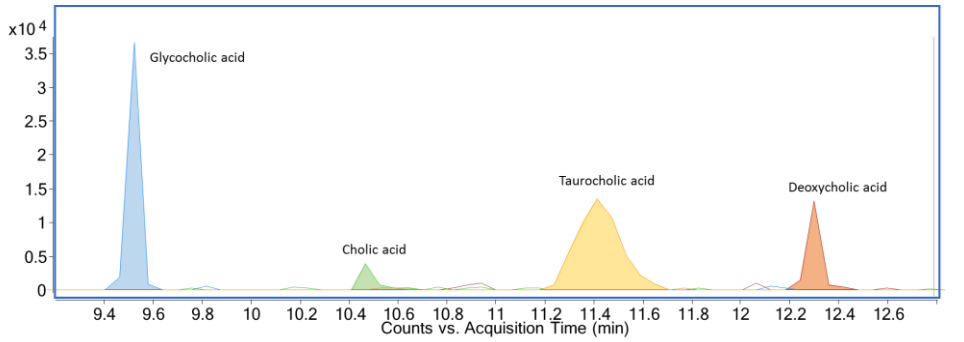


Figure S-5. Extracted ion chromatograms (EICs) of identified bile acids in negative ionization mode: glycocholic acid (464.3039 *m/z*), cholic acid (407.2812 *m/z*), taurocholic acid (514.2877 *m/z*), and deoxycholic acid (391.2877 *m/z*).

Table S-1. List of all metabolites identified in serum classified by families and including the information that led to their identification: the MS and MS/MS information (only two of the product ions are shown), as well as the retention times.

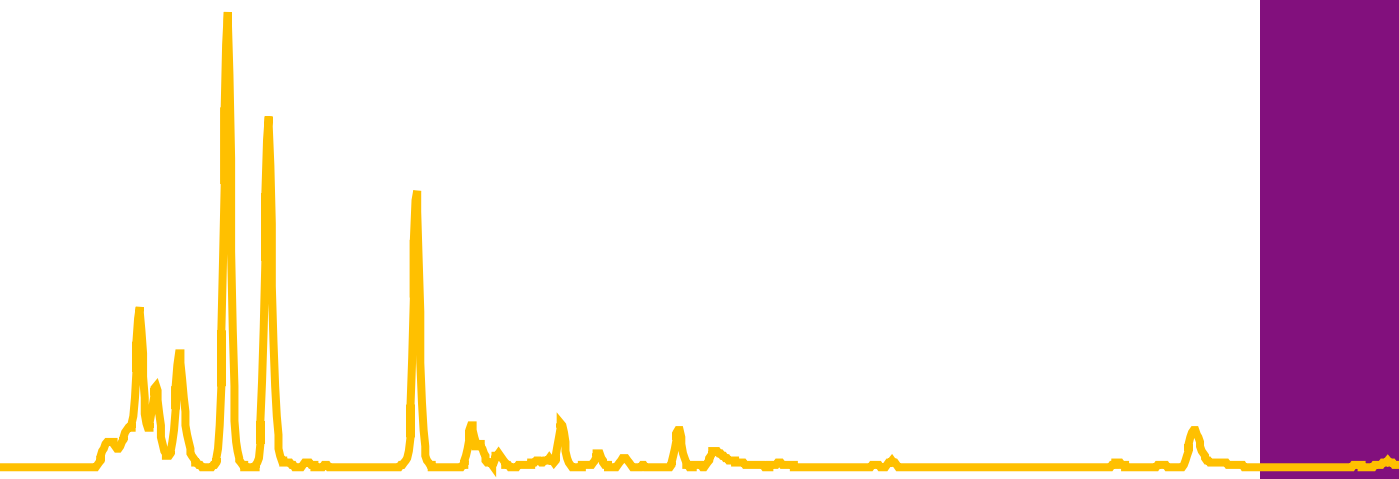
Family	Compound	Precursor ion (*m/z*)	Adduct	Product ions (*m/z*)		R_t (min)
Acylcarnitines	Acetylcarnitine	204.1234	$[M+H]^+$	145.0493	85.0287	1.0
	Isobutyrylcarnitine	232.1536	$[M+H]^+$	173.0758	144.1036	5.4
	Octanoylcarnitine	288.2164	$[M+H]^+$	229.1509	144.1036	7.9
	Decanoylcarnitine	316.2486	$[M+H]^+$	155.1428	257.1738	8.9
Amino acids and derivates	Betaine	118.0867	$[M+H]^+$	58.0658	59.0735	0.7
	Carnitine	162.1124	$[M+H]^+$	103.0387	-	0.8
	Creatine	132.0769	$[M+H]^+$	114.0635	90.0540	0.8
	Leucine/isoleucine	132.1016	$[M+H]^+$	86.0960	-	2.0
	Leucine/isoleucine	132.1016	$[M+H]^+$	86.0960	-	2.2
	Tyrosine	182.0812	$[M+H]^+$	165.0557	136.0776	2.5
	Phenylalanine	166.0856	$[M+H]^+$	120.0803	103.0545	5.2
	Phenylglycine	152.0706	$[M+H]^+$	107.0477	135.0371	5.4
	Tryptophan	205.0964	$[M+H]^+$	188.0703	146.0596	5.9
	Hippuric acid	178.0510	$[M-H]^-$	134.0576	77.0410	6.9
	Indoxylsulfuric acid	212.0025	$[M-H]^-$	132.0440	79.9582	8.7
Bile acids	Glycocholic acid	464.3039	$[M-H]^-$	402.2986	74.0210	9.5
	Cholic acid	407.2812	$[M-H]^-$	241.0738	-	10.5
	Taurocholic acid	514.2877	$[M-H]^-$	124.0242	-	11.2
	Deoxycholic acid	391.2877	$[M-H]^-$	345.2739	-	12.3

Dicarboxilic acids	Citric acid	191.0201	$[M-H]^-$	111.0108	87.0032	1.9
	Nonanedioic acid	187.0977	$[M-H]^-$	169.0871	125.1049	8.2
Fatty acids and derivates	Leukotriene B4 (LTB4)	335.2244	$[M-H]^-$	195.0960	129.0533	11.7
	13-HpODE	311.2246	$[M-H]^-$	293.2173	113.0918	13.0
	9-HODE	295.2305	$[M-H]^-$	277.2173	171.1036	13.1
	12-HETE	319.2295	$[M-H]^-$	301.2180	135.1174	13.7
	Linolenic acid (C18:3)	277.2173	$[M-H]^-$	259.2017	-	15.2
	Docosahexaenoic acid (C22:6)	327.2354	$[M-H]^-$	283.2446	229.1943	15.3
	Arachidonic acid (C20:4)	303.2336	$[M-H]^-$	259.2481	205.1933	15.5
	Linoleic acid (C18:2)	279.2354	$[M-H]^-$	-	-	15.8
	Oleic acid (C18:1)	281.2517	$[M-H]^-$	-	-	16.5
Lysophospholipids	LPC(20:5)	542.3241	$[M+H]^+$	184.0708	104.1025	11.1
	LPC(18:3)	518.3206	$[M+H]^+$	184.0772	104.1068	11.2
	LPC(16:1)	494.3241	$[M+H]^+$	184.0735	104.1060	11.4
	LPE(18:2)	478.2922	$[M+H]^+$	337.2728	-	11.5
	LPE(20:4)	502.2924	$[M+H]^+$	361.2726	-	11.6
	LPE(22:6)	526.2917	$[M+H]^+$	385.2662	-	11.7
	LPC(22:6)	568.3398	$[M+H]^+$	184.0747	104.1148	11.8
	LPC(18:2)	520.3365	$[M+H]^+$	184.0727	104.1045	11.9
	LPC(20:4)	544.3376	$[M+H]^+$	184.0683	104.1075	12.0
	LPE(16:0)	454.2926	$[M+H]^+$	313.2745	-	12.0
	LPC(22:5)	570.3551	$[M+H]^+$	184.0693	104.1049	12.1
	LPC(20:3)	546.3525	$[M+H]^+$	184.0743	104.1097	12.2
	LPE(20:3)	504.3102	$[M+H]^+$	363.2936	-	12.2
	LPC(16:0)	496.3397	$[M+H]^+$	184.0668	104.1048	12.4
	LPE(18:1)	480.3080	$[M+H]^+$	339.2881	-	12.5
	LPC(O-16:1)	480.3448	$[M+H]^+$	184.0681	104.1032	12.6
	LPC(18:1)	522.3549	$[M+H]^+$	184.0730	104.1048	12.6
	LPC(22:4)	572.3689	$[M+H]^+$	184.0760	104.1032	12.8
	LPC(O-16:0)	482.3604	$[M+H]^+$	184.0762	104.1064	12.8
	LPC(17:1)	508.3751	$[M+H]^+$	184.0724	104.1064	13.0
	LPE(18:0)	482.3237	$[M+H]^+$	341.3035	-	13.6
	LPC(18:0)	524.3716	$[M+H]^+$	184.0723	104.1068	13.8
Phospholipids	PC(18:2/20:4)	850.5559	$[M+HCOO]^-$	790.5390	303.2291	14.5
	PC(18:1/18:2)	828.5747	$[M+HCOO]^-$	768.5515	281.2811	14.7
	PC(19:1/20:2)	870.5686	$[M+HCOO]^-$	810.5444	295.2865	14.7
	PC(16:0/18:2)	802.5586	$[M+HCOO]^-$	255.2295	279.2294	14.8
	PC(18:1/18:1)	830.5884	$[M+HCOO]^-$	770.5730	281.2818	14.9
	PC(18:1/20:4)	852.5688	$[M+HCOO]^-$	792.5429	281.2830	14.9
	PC(18:0/18:1)	832.6003	$[M+HCOO]^-$	772.5957	281.2821	15.0
	PC(16:0/18:1)	804.5804	$[M+HCOO]^-$	255.2350	281.2434	15.0
	PC(16:0/19:1)	818.5521	$[M+HCOO]^-$	758.5389	295.2645	15.1

Purines and derivates	Purine	121.0509	$[M+H]^+$	113.9757	104.0199	1.3
	Uric acid	169.0354	$[M+H]^+$	152.0090	141.0371	2.1
	Xanthine	151.0266	$[M-H]^-$	108.0332	-	2.5
	Hypoxanthine	137.0457	$[M+H]^+$	119.0378	110.0465	2.6
Sphingolipids	Sphingosine	300.2895	$[M+H]^+$	282.2781	252.2649	9.8
	Sphingosine-1-phosphate	380.2569	$[M+H]^+$	282.2733	264.2723	10.9
Vitamins	Choline	104.1074	$[M+H]^+$	60.0803	-	0.7

SECTION II

Profiling analysis as a tool for biomarkers searching and nutrimetabolomics studies



Section II of this PhD Book is devoted to two research areas with high development in clinical metabolomics at present: nutrimetabolomics and biomarkers search. The research on both is under the umbrella of profiling analysis.

In dealing with nutrimetabolomics, to which the research in Chapter 4 is dedicated, metabolome alterations caused in individuals subjected to four intervention diets demonstrate, once again, the importance of diet on health. The role of this omics on the search for a personalized diet is one of the key steps towards a personalized medicine.

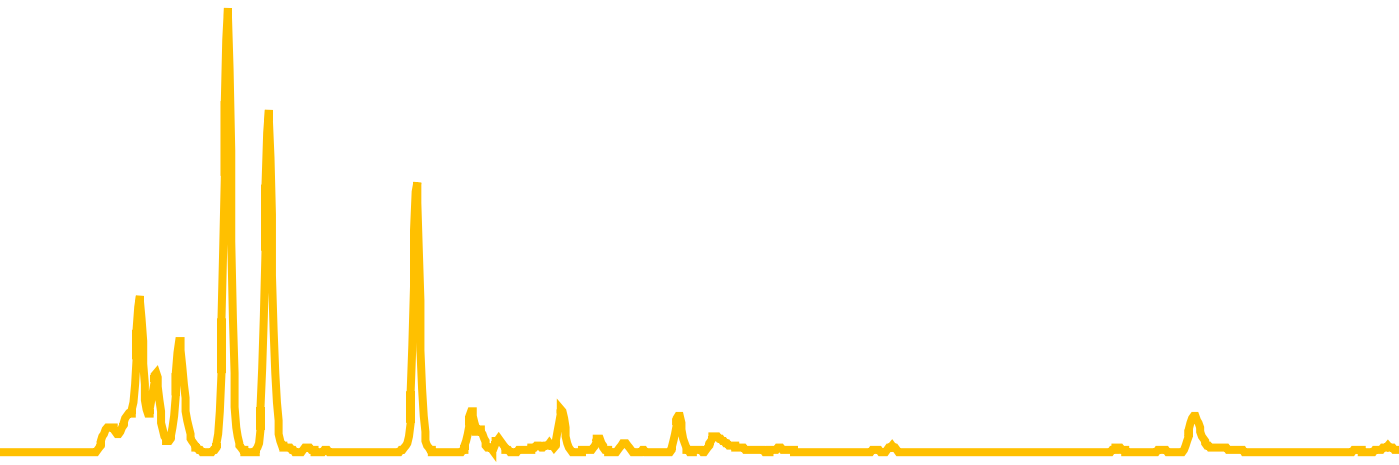
Two types of samples (*serum* and *sweat*) from patients with two very different types of diseases (*atherosclerosis* and *lung cancer*) were the basis for the developed research together with proper analytical approaches such as LC–QTOF MS/MS, sample preparation, if required, and application of the most convenient data treatment.

Serum from patients diagnosed with *atherosclerosis* and affected by stable angina or myocardial infarction (acute myocardial infarction or non-ST elevation myocardial infarction) were the target samples to be analyzed in Chapter 5 and 6. The profiles of given serum metabolites were compared to search for potential biomarkers for early prediction of an ischemic event, which could lead to a drastic reduction of mortality. Change in the concentration of statistically significant metabolites allowed discrimination of individuals who suffered acute or non-ST elevation myocardial infarction from those suffering stable angina, as can be seen in Chapter 5. On the other hand, a three-variable panel for infarct prediction was created, thus discriminating between individuals affected by acute myocardial infarct and those diagnosed by stable angina with sensitivity and specificity values higher than 80%, as shown in Chapter 6.

Sweat from patients with *lung cancer* and control individuals were analyzed by the analytical platform in Chapter 1, and the individual capability of the identified metabolites to discriminate between the two groups in terms of sensitivity and specificity was successfully established. Additional individual or combined panels of metabolites allowed reducing false negative and false positive rates, as discussed in Chapter 7.

Chapter 4:

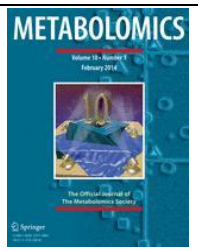
Comparative nutrimetabolomics study of the influence of fat intervention diets on serum metabolic profiles by LC–MS/MS



Springer

Sent to

Metabolomics



Comparative nutrimetabolomics study of the influence of fat intervention diets on serum metabolic profiles by LC–MS/MS

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

Comparative nutrimetabolomic study of the influence of fat intervention diets on serum metabolic profiles by LC–MS/MS

*M. Calderón-Santiago, F. Priego-Capote, M. D. Luque de Castro**

Abstract

A nutrimetabolomic study to assess the influence of fat intervention diets on serum metabolic profiles in 76 patients diagnosed with metabolic syndrome is here presented. Four isoenergetic diets but with different fat content [two with high-fat content (38%) and two with low-fat content (28%)] were administered to the selected cohort. Metabolite profiles were obtained before and after the intervention period for 12 weeks by serum fractionation into polar and non-polar metabolites and LC–QTOF analysis in high resolution mode. Multivariate statistical analysis allowed discrimination among the post-intervention samples and between pre- and post-intervention stages for the four diets. Within the lipid fraction, phospholipids were the family of metabolites more influenced by the intervention diets. Concretely, three common glycerophospholipids [PC(18:0/16:2), PC(16:0/18:2) and PC(18:0/18:2)] were affected by the diets. High-fat diets led to 19 altered phospholipids, while 16 phospholipids showed changes by low-fat diets. Concerning the polar fraction, high-fat diets altered more metabolites (14–15) than low-fat diets (6–9). In this fraction, amino acids were mostly influenced after the high-fat diets, and only ornithine was altered with one of the low-fat diets. Bile acids and acylcarnitines levels were also altered after diet intake. A common behavior was found for some of the identified compounds; thus, octenoyl-L-carnitine and lysophosphatidylcholine (16:0) levels appeared increased after intake of any of the diets.

1. Introduction

Nutritional studies play a key role in the identification of essential nutrients needed for human growth and health. Nowadays, nutritional scientists are challenged to find new ways to treat or prevent diseases brought on by nutritional oversufficiency such as obesity, diabetes, chronic inflammation and/or cardiovascular diseases. One of the main approaches to face up these studies is nutrimetabolomics, which is aimed at: (1) food component analysis; (2) food quality/authenticity detection; (3) consumption biomarkers; and, (4) elucidation and detection of effects and pathology-risk biomarkers in nutritional and epidemiological studies. The last approach has been used to assess metabolic changes caused by intake of certain nutrients with an impact on human health. The starting hypothesis is that the majority of metabolism alterations are detected in biological fluids such as serum/plasma or urine. One of the most common analytical strategies to detect and discriminate metabolic changes is metabolomics fingerprinting, which is aimed at developing classification models based on snapshots or fingerprints from samples by suited detection techniques [1]. However, data quality in these approaches is limited as low discrimination among metabolites is attained. Data quality is drastically enhanced by application of a global metabolomics approach focused on detection and identification of those metabolites with the highest contribution to the observed variability. This strategy seems to be especially suited to detect and discriminate metabolic changes caused by nutritional states.

Considering the chemical diversity and number of metabolites in human biofluids (a conservative estimation of human serum reported 4229 metabolites and a wide range of concentrations —up to 11 orders of magnitude— observed in human serum [2–4]), different analytical platforms in combination are needed to approach full coverage of the metabolome. In fact, a variety of techniques can be employed to maximize metabolite detection coverage, with nuclear magnetic resonance (NMR) and mass spectrometry (MS) being the most used. The coupling of MS to either LC (LC–MS) or gas chromatography (GC–MS) has been widely used to study metabolic changes associated with pathological states and

nutritional studies [5–10]. The choice of a separation technique prior to MS detection is primarily dictated by the properties of the given analyte, with GC–MS as more suitable for volatile and semi-volatile analytes and capillary electrophoresis (CE) as an alternative to LC for charged analytes [11,12].

Comparing the use of LC–MS to NMR for metabolomics studies [7,13,14], the sensitivity of the former is significantly higher than that characterizing NMR, thus providing the possibility to monitor metabolic changes that remained unnoticed when working with the latter. LC–MS allows analysis of a wide range of metabolite classes. The widespread availability of robust LC–MS methods in many research laboratories has increased enormously its use for metabolic profiling over the past few years. In fact, many studies have provided higher metabolome coverage (approximately 55–70% of total metabolites detected using multiple platforms) by a single experiment using LC–MS than by GC–MS or CE–MS [4] (an estimated 10–15% overlap is usual between LC–MS and GC–MS platforms [8]). Therefore, LC–MS can provide comprehensive metabolite coverage by untargeted analyses using a single platform [15]. LC–MS metabolomics studies generally involve differential comparison of several subjects or treatment groups (*e.g.* normal *versus* treated individuals, healthy *versus* ill individuals) without *a priori* identification of all detected metabolites.

The Lipgene project is a large European multi-centre project which includes a human intervention trial (486 volunteers at baseline) conducted to compare the impact of different types and amounts of dietary fatty acids on insulin sensitivity. This human intervention study is supported on a food-exchange model involving four diets for alteration of dietary fat quantity and quality in participants. The key aspects of this strategy for correlation and discrimination of metabolic patterns with the supplied diets were: (i) diets differ significantly in overall fat intake but remain isoenergetic; (ii) the two high-fat diets are significantly higher in the content of saturated fatty acids (SFAs) or monounsaturated fatty acids (MUFAs) as compared with the two low-fat diets; (iii) the effect of long-chain *n-3* polyunsaturated fatty acids (PUFAs) supplementation was tested in a low-fat background diet. Supported on this

study, an LC–QTOF MS/MS platform has been applied to obtain metabolic profiles of serum from individuals subjected to four isoenergetic intervention diets under controlled conditions for twelve weeks. The aim was to detect and evaluate serum metabolic differences according to the administered diets.

2. Materials and methods

2.1. Chemicals

LC-MS grade acetonitrile (Sigma-Aldrich, Madrid, Spain) was used for preparation of chromatographic mobile phases. MS-grade formic acid (Scharlab) was used as ionization agent in LC-QTOF analyses, while deionized water (18 MΩ·cm) from a Millipore Milli-Q water purification system (Millipore, Bedford, MA, USA) was used to prepare the chromatographic aqueous phases. Chromatographic-grade methanol and chloroform (Scharlab, Barcelona, Spain) were used for sample treatment.

2.2. Instruments and apparatus

A Sorvall Legend Micro 21R centrifuge (Thermo Scientific, Waltham, MA, USA) was used to centrifuge samples after deproteination. An Agilent 1200 Series LC system (consisting of a binary pump, a vacuum degasser, an autosampler and a thermostated column compartment) coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with a dual electrospray (ESI) source (Santa Clara, CA, USA) was used. The chromatographic eluate was monitored in high-resolution mode.

2.3. Diets composition

Four isoenergetic diets that differed in fat quantity and quality were planned for this study. Two diets were designed to provide 38% energy (E) from fat and the other two with 28% E from fat with the following assigned names and compositions:

(1) HSFA: high-fat content (38% energy) with high proportion of saturated fatty acids (16% SFA), while monounsaturated and polyunsaturated fatty acids (MUFAs and PUFAs, respectively) were present at 12% and 6%, respectively.

(2) HMUFA: high-fat content (38% energy) with high proportion of MUFAs (20%), while SFA and PUFAs were 8% and 6%, respectively.

(3) LFHCC: low-fat content (28% energy) with high-complex carbohydrate diet (8% SFA, 11% MUFA, 6% PUFA), supplemented with 1 g/d high oleid acid sunflower oil supplement.

(4) LFHCCn-3: low-fat (28% energy), high-complex carbohydrate diet (8% SFA, 11% MUFA, 6% PUFA), supplemented with 1.24 g/d very long chain *n-3* PUFA supplement.

2.4. Cohort under study

This research was conducted within the framework of the LIPGENE integrated project [diet, genomics and metabolic syndrome (MetS): an integrated nutrition, agro-food, social and economic analysis]. A total of 76 patients with MetS from the LIPGENE cohort were accepted to participate in this study aimed at evaluating metabolic alterations by LC-QTOF MS/MS. All participants gave written informed consent and underwent a comprehensive medical history, physical examination and clinical chemistry analysis before enrolment. None of the subjects was taking medication or supplementary vitamins with influential effect on serum metabolome.

Patients were randomly divided into four groups to receive one of the four dietary interventions for 12 weeks under controlled conditions. Both the design and protocol of the intervention study have been described in detail by Shaw *et al.* [16]. Serum control samples were obtained in basal state prior to the intervention as control samples (PRE). Patients arrived at clinical centres after 12 weeks of the intervention study following a 12-h fast refrained from smoking during the fasting period and abstained from alcohol intake during the preceding

7 days. In the laboratory and after cannulation a fasting blood sample was taken for each individual (POS).

2.5. Blood extraction and serum isolation

Sampling was performed following the Recommendations on Biobanking Procedures for serum processing and management recently published by the European Consensus Expert Group [17]. The subsequent aspects were taken into account: (i) sample storage at 80°C; (ii) recording of the time from collection through processing; (iii) experimental definition of the time limits appropriate for analytes measurement. All steps from blood to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004), which were supervised by specialized personnel.

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated for 30 min at room temperature to allow coagulation. Then, the tubes were centrifuged at 2000 × *g* for 15 min at 4 °C to isolate the serum fraction that was placed in plastic tubes and stored at –80 °C until analysis.

2.6. Sample treatment

Serum samples (100 µl) immersed in a bath ice were treated with 300 µl of 1:2 methanol–chloroform. The mixture was shaken for 2 min and centrifuged for 5 min at 4 °C and 20200 × *g*. Both phases (aqueous and organic) were collected in different vials, then placed in the LC autosampler for subsequent analysis.

2.7. LC–QTOF MS/MS analysis

Chromatographic separation was performed using a C18 reverse-phase analytical column (Mediterranean, 100 mm x 0.46 mm i.d., 3 µm particle size) from Teknokroma (Barcelona, Spain), which was thermostated at 25°C. The mobile phases were 5% ACN (phase A) and 95% ACN (phase B) both with 0.1% formic acid as ionization agent. Two chromatographic modes were programmed according to the two phases obtained from the sample. In both cases the LC

pump was programmed with a flow rate of 0.8 mL/min. For the aqueous fraction, the following elution gradient was set: 9% phase B was fixed as initial mobile phase, increased to 100% of phase B from min 0 to 10, and kept at 100% for 5 min. On the other hand, the elution gradient established for the non-polar fraction was: 40% phase B was fixed as initial mobile phase, increased to 100% of phase B from min 0 to 10, and kept at 100% for 5 min. In both cases, a post-time of 5 min was set to equilibrate the initial conditions for the next analysis. The injection volume was 8 µL and the injector needle was washed for 10 times between injections with 80% methanol. Furthermore, the needle seat back was flushed for 12 s at a flow rate of 4 mL/min with 80% methanol to avoid cross contamination.

The parameters of the electrospray ionization source, operating in negative and positive ionization mode (negative mode for non-polar fraction and positive mode for polar fraction), were as follows: the capillary and fragmentor voltage were set at ±3.5 kV and 110 V, respectively; N_2 in the nebulizer was flowed at 40 psi; the flow rate and temperature of the N_2 as drying gas were 10 L/min and 350°C, respectively. The instrument was calibrated and tuned according to procedures recommended by the manufacturer. MS and MS/MS data were collected in both polarities using the centroid mode at a rate of 2.5 spectrum per second in the extended dynamic range mode (2 Ghz). Accurate mass spectra in auto MS/MS mode were acquired in MS m/z range 100–1100 and MS/MS *m/z* range 31–1100. The auto MS/MS mode was configured with 2 maximum precursors per cycle and an exclusion window of 0.25 min after two consecutive selections of the same precursor. The collision energy selected was 20 eV.

The instrument gave typical resolution 15000 FWHM (Full Width at Half Maximum) at m/z 118.086255 and 30000 FWHM at m/z 922.009798. To assure the desired mass accuracy of recorded ions, continuous internal calibration was performed during analyses by using the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated hexakis (1H, 1H, 3H-tetrafluoropropoxy) phosphazine or HP-921] in positive ion mode; while in negative ion mode ions

with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct of HP-921) were used.

2.8. Data processing and statistical analysis

MassHunter Workstation software package (B.05.00 Qualitative Analysis and B.06.00 Profinder, Agilent Technologies, Santa Clara, CA, USA) was used to process all data obtained by LC-QTOF in auto MS/MS mode. The MassHunter Profinder software was employed to extract and align potential molecular features in all the injections by using the recursive feature extraction algorithm. This algorithm performs firstly a chromatographic deconvolution and aligns the features across the selected sample files using mass and retention time. Secondly, it uses the mass and retention time of each feature to perform a recursive targeted feature extraction. This two-step procedure reduces the number of false negatives as well as false positives in feature extraction. Parameters considered for feature extraction included a threshold of 1500 counts and a maximum charge state of two. Additionally, the isotopic distribution for a 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). Adducts formation in the positive (+H, +Na) and negative ionization (-H, +HCOO) modes were also included to identify features corresponding to the same molecule. Features alignment was done with a tolerance window of 0.45 min and 10 ppm mass accuracy for retention time and *m/z* values across all the data files, respectively. The minimum peak area required for the feature extraction was fix on 3000 counts, and this value was also set at 3000 counts for at least the 75 percent of the samples of each group considered in the cohort in the recursive step.

After features extraction and alignment, data files in compound exchange format (.cef files) were created for each sample and exported into the Mass Profiler Professional (MPP) software package (version 12.1, Agilent Technologies, Santa Clara, CA, USA) for further processing. Data pretreatment was based on baselining to remove background noise and normalization by logarithmic transformation to reduce relatively large differences among molecular feature

abundances. Supplementary Figure 1 shows the normalization effect of the logarithmic transformation. As can be seen, the resultant matrix fits a normal distribution both in the polar and non-polar fraction.

MPP software allows both supervised and unsupervised analysis of the data by PLS-DA and PCA. In the former case, the validation model selected was N-Fold, by which the classes in the input data were randomly divided into N equal times; N-1 parts were used for training, and the remaining one part was used for testing. The process was repeated N times, with a different part used for testing in each iterative step. Then, repetitions and a fold number of 3 were selected for all validations.

Once the difference between the groups under study were verified, an analysis of variance (ANOVA) was executed using the ROCCET online tool (http://www.roccet.ca/ROCCET/) to know the entities significantly differenced between sampling times PRE and POS in each diet. The full process was done separately with each fraction: aqueous and organic phases.

3. Results and discussion

3.1. Serum metabolite differences according to the intervention diets

A common factor associated to metabolomics analysis in clinical and nutritional studies dealing with humans is the biological variability among individuals, which frequently mask the effects of other variability sources associated to internal factors (diseases, metabolic disorders) or external factors (diet, lifestyle). In this research, data acquisition was carried out to observe the biological variability not ascribed to diets intake as compared to the variability after diets intake. For this purpose, the chromatograms acquired in basal state (PRE) were compared with those acquired after the intervention period for 12 weeks (POS) of each diet. After obtaining a global profile of metabolites present in polar and non-polar phases representing individuals subjected to every diet, the comparison among the different profiles would allow evaluation of

similarity/dissimilarity patterns according to administered diets. Supplementary Figure 2 shows the base peak chromatograms (BPCs) obtained in both ionization modes from an individual prior the intervention study (PRE) and after the intervention diets (POS).

As mentioned under 'Materials and methods', only those molecular entities detected in at least 75% of the samples belonging to one of the groups considered under study (PRE and POS for each diet) were considered. This procedure allows reducing the presence of tentative metabolites related to inter-individual variability sources such as anthropometric factors. With these premises, the number of molecular features detected in the non-polar fraction (negative ionization mode) was 170, while in the polar fraction (positive ionization mode) 41 tentative compounds were detected.

After data pretreatment, Principal Component Analysis (PCA) was employed in the first stage of data treatment to evaluate if the variability associated to the intervention study allowed discriminating POS samples from PRE samples. For this purpose, samples in the PCA scores plot were labelled by the sampling time but no attention was paid to the type of diet. Supplementary Figure 3 shows the PCA scores plot in a 3D graph. As can be seen, no clear discrimination was observed between PRE and POS states although certain trends can be visualized. The variability explained with PC1, PC2 and PC3 was 33.9% for the organic phase and 27.3% for the aqueous phase, which clearly means that the variability associated to the intervention diets was not the major cause of variability present in the samples cohort. Taking into account this fact, a PLS-DA was carried out to find discrimination according to the intervention diet. In this case, five classes were considered: PRE samples and POS samples for the four intervention diets. Figure 1 shows the PLS-DA loadings plot obtained both for polar and non-polar fractions, which clearly allowed differentiating individuals prior to the intervention diets (PRE sampling time) from individuals after the intervention period (POS sampling time). Additionally, no complete discrimination was observed among individuals subjected to the four diets. Nevertheless, certain groupings were detected in both fractions according to the

intervention diet and, additionally, individuals after the low-fat diet supplemented with long chain *n-3* PUFAs (LFHCCn-3) provided the highest variability versus control individuals (PRE sampling time) in comparison to the resting diets. If the PRE sampling time is not considered for PLS-DA, separation of individuals after the four intervention diets was appreciated, as Figure 2 shows. As can be seen, discrimination of individuals according to the intervention diet was observed in both the polar and the non-polar fractions. In both analyses the maximum separation was observed for individuals subjected to the LFHCCn-3 diet, who were assigned with the most discriminant metabolic pattern. Concerning the diet influence, the main variability source was not associated to the fat content but the diet composition. No similarity groupings were observed between pairs of diets. At this point, the effects found for the polar and non-polar fractions as a result of the intervention diets were independently evaluated

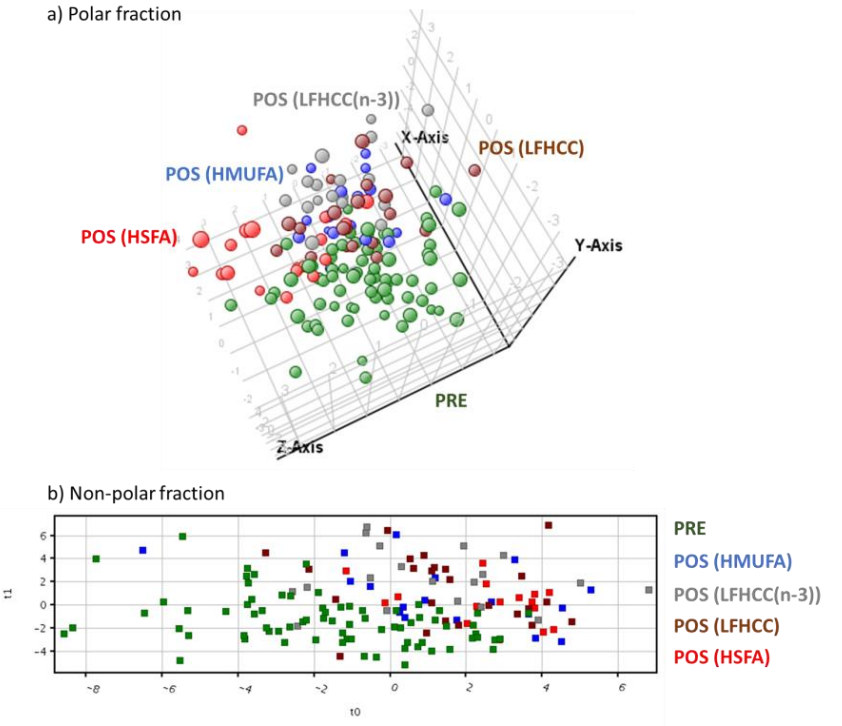


Figure 1. Scores graphs obtained after multivariate analysis by PLS-DA of the polar and non-polar fraction, considering POS and PRE time for each diet.

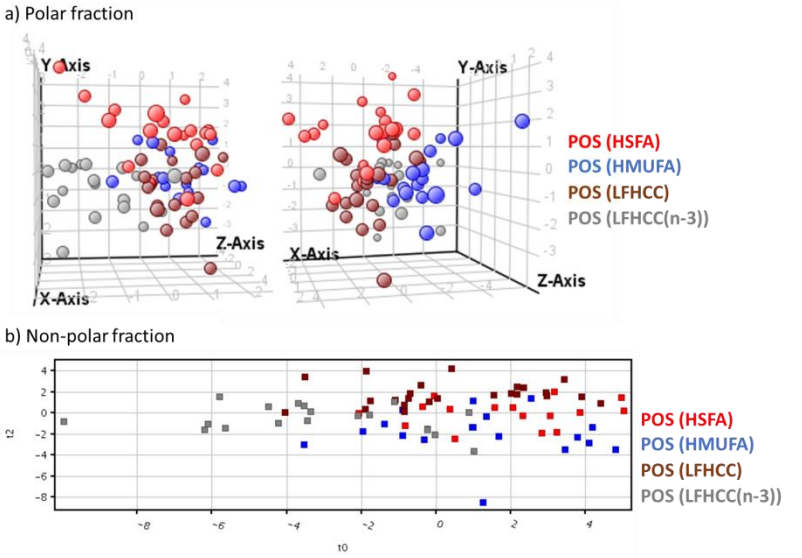


Figure 2. Scores graphs obtained after multivariate analysis by PLS-DA of the polar and non-polar fractions, considering POS time for each diet.

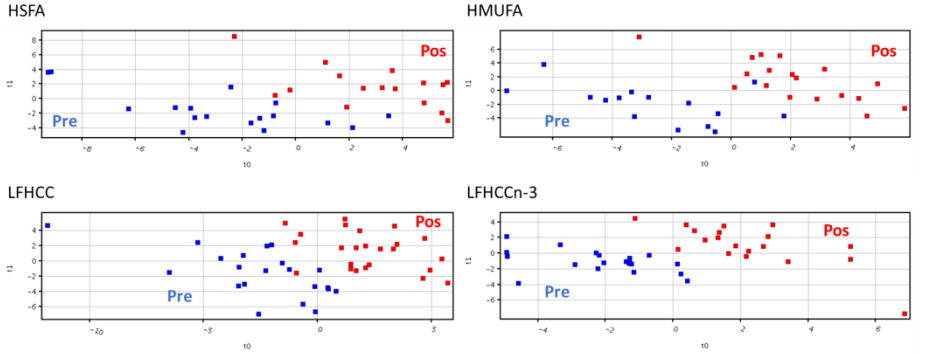


Figure 3. Scores graphs obtained after multivariate analysis by PLS-DA of non-polar fraction for each diet, considering in each case PRE and POS times.

3.2. Non-polar metabolic effects associated to the intervention diets

The influence of each diet on non-polar metabolites in serum was independently evaluated by comparing the PRE sampling time versus the POS sampling time for each diet. For this purpose, a PLS-DA model was created for each diet as Figure 3 shows. Complete discrimination was observed between

individuals before and after the intervention diets. The discrimination capability obtained in the training step was above 94% for the four diets.

Taking into account the existing metabolic differences caused by each intervention diet, a fold change analysis was executed to identify the tentative metabolites with the highest contribution capability to explain the observed variability. In all cases, a number of molecular features between 17 and 19 experienced a fold change higher than 1.25, which implies an increase or decrease of 25% of the MS detector response in relative terms. Identification of these tentative metabolites by using MS and MS/MS allowed detecting a significant number of phospholipids, essentially, glycerophosphatidylcholines and glycerophosphatidylethanolamines including different fatty acids chains. Table 1 lists the lipids identified after fold change analysis. The significance level according to an ANOVA test is also shown. As can be seen metabolic changes occurring in the non-polar fraction of serum after four intervention diets for 12 weeks were targeted at phospholipids. Apart from these two families of phospholipids, several sphingomyelins and two essential fatty acids —linoleic acid and docosahexanoic acid— were also detected with a concentration change due to intervention diets. Table 1 shows all the non-polar metabolites that experienced a fold change of at least 1.25 by comparing physiological states before or after the intervention diets. They are grouped as a function of the diet but also according to the magnitude of change. The first aspect to be emphasized is regarding the number of glycerophospholipids or sphingomyelins with two substituents that experienced a concentration change above 1.25 (increase or decrease in individuals after the intervention diets). Thus, two diets, HSFA and LFHCC, presented increased levels of 7 and 8 glycerophospholipids, and sphingomyelins with two substituents, respectively. On the other hand, only three or one metabolites of these families decreased their concentration due to the intake of these two intervention diets. The opposite behavior was observe for the LFHCCn-3 diet since only three of these compounds increased their concentration in serum due to the intervention, while 7 of them decreased their concentration. The fourth diet —HMUFA diet— caused an increase of four disubstituted

glycerophospholipids or sphingomyelins and a decrease of the same number of lipidic metabolites. Therefore, the effect of the intervention diet on the lipid profile, which was mostly manifested on phospholipidic structures, depended on the diet. Nevertheless, some common effects were observed in individuals after any intervention diet. The most representative example was found for lysophospholipids, with special emphasis on lysoPC(O-16:1), that experienced an increase of serum concentration in all individuals (>2.0) after the intervention diets. In fact, this change was statistically significant according to the ANOVA test for individuals subjected to HSFA, LFHCC and LFHCCn-3 diets.

Table 1. Non-polar metabolites that experienced a fold change above 1.25 by comparing physiological states before or after intervention diets. Isomers of the same lysophosphatidylcholine are differenced by (1) and (2).

	HSFA	HMUFA	LFHCC	LFHCC(n-3)
>2	SM(d22:0/20:3)** LysoPC(O-16:1)** PC(16:0/18:2) LysoPE(O-16:1) PC(18:0/16:2) PC(O-18:0/18:2)*	LysoPC(18:2) (1) LysoPC(18:2) (2) LysoPC(16:0) (2) LysoPC(O-16:1)	PC(16:0/18:2)*** PC(18:0/18:2)* SM(d20:1/16:1) LysoPC(O-16:1)*** PC(38:5)[1] PC(18:0/20:4)	LysoPC(O-16:1)**
1.75-2	LysoPE(18:0) LysoPC(18:0) (1) PC(18:0/18:2) SM(d18:2/24:0)	LysoPC(16:1) SM(d20:1/16:1) SM(d22:0/20:3)	LysoPC(18:0) (1)	PC(16:0/18:2) Docosahexaenoic acid (C22:6)
1.5-1.75		PC(18:0/20:4) PC(16:0/20:3)	LysoPC(17:0) PC(18:0/16:2)	PC(O-18:0/18:2)
1.25-1.5	LysoPC(16:0) (1) LysoPC(16:0) (2) PC(18:2-OH/16:0) LysoPE(18:1) LysoPC(18:2) (2)	LysoPC(18:0) (1) LysoPC(18:0) (2)	LysoPC(16:0) (1) PC(16:0/20:3) SM(d18:0/18:1)	LysoPC(16:0) (1) SM(d20:1/16:1) LysoPC(20:4)
(-1.25)-(-1.5)		Docosahexahenoic acid (C22:6) PC(18:0/16:2) LysoPE(18:1)	Linoleic acid (C18:2)	PC(18:0/18:1)
(-1.5)-(-1.75)				PC(18:0/18:2) PC(38:5)[1]
(-1.75)-(-2)	PC(22:6/18:0) PC(O-15:1/18:1) LysoPC(18:2) (1)			PC(16:0/20:3)
<(-2)	PC(16:0/20:4)	PC(18:0/18:2)*** PC(16:0/18:2) LysoPC(16:0) (1) PC(O-18:0/18:2) LysoPC(17:0)	PC(16:0/20:4) LysoPC(20:4) LysoPE(20:4)	LysoPC(18:2) (2)** PC(18:2-OH/16:0) LysoPE(20:4) SM(d22:0/20:3) PC(18:0/16:2)

*** p-value < 0 .01
** 0.01<p-value<0.025
* 0.025<p-value
[1] PC(38:5) is the sum of PC(18:0/20:5) and PC(16:0/22:5)

Apart from these overall changes, some particular modifications in families of phospholipids were observed as a function of the diet. Thus, the main difference among diets was observed for three of the most common glycerophospholipids, PC(16:0/18:2), PC(18:0/16:2) and PC(18:0/18:2). These three metabolites reported different levels depending on the diet, as Figure 4 shows. Thus, the concentrations of these three PCs were increased in individuals after HSFA and LFHCC interventions, while the same compounds were decreased in individuals subjected to HMUFA diet. On the other hand, individuals after the LFHCCn-3 intervention diet were marked by a different effect since PC(16:0/18:2) increased its relative concentration in serum, while PC(18:0/18:2) and PC(18:0/16:2) levels were decreased. Attending to these results, a direct influence of the intervention diet on three of the most important blood phospholipids was found.

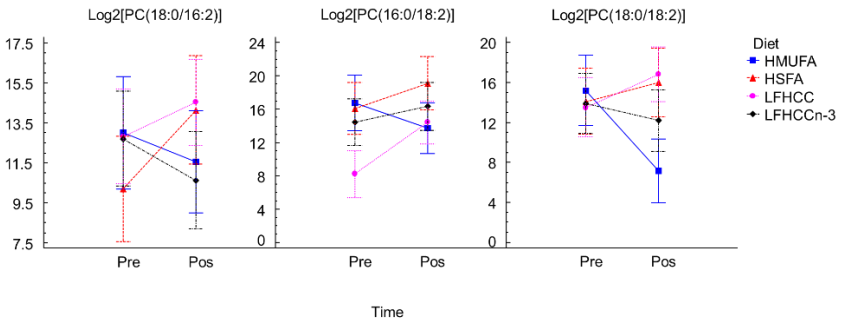


Figure 4. Mean values and 95% confidence levels of the normalized area for PC(18:0/16:2), PC(16:0/18:2) and PC(16:0/18:2).

Additional effects associated to the intervention diets can be deduced by comparing the compounds that change their concentration in relative terms. Thus, the two diets characterized by high-fat content (38% energy) led to an opposite effect in the concentration of glycerophospholipids combining essential PUFA substituents and relevant saturated FAs such as palmitic and stearic acids (C16:0 and C18:0). Thus, PC(22:6/18:0) and PC(16:0/20:4), which include docosahexaenoic acid and araquidonic acid, decreased their concentration in serum after intake of the HSFA diet. On the other hand, PC(18:0/20:4) and

PC(16:0/20:3) increased their levels in serum after the HMUFA diet. Concerning the lysophospholipids profile, these two diets described a common pattern since a quite similar trend was found for the most important lysophospholipids — particularly, lysoPC(16:0) and lysoPC(18:0)— except for one of the isomers of lysoPC(18:2) that described an opposite effect. This metabolite (eluting at 8.0 min) increased its concentration in HMUFA individuals, and decreased in individuals subjected to HSFA diet. One other difference detected in individuals controlled by these two diets was that affecting lysoPE(18:1). This compound increased its concentration in HSFA individuals, while this was decreased in HMUFA cases.

Concerning the two diets based on low-fat content (LFHCC and LFHCCn-3), a similar trend was observed for most phospholipid structures, except for glycerophospholipids. Glycerophosphocholines, including PUFA substituents, clearly increased their concentration in LFHCC individuals, while LFHCCn-3 individuals were marked by a decreased of this PC structure. Also, one other difference between individuals subjected to these two diets was observed for lysoPC(20:4), which includes arachidonic acid. This lipidic metabolite was found at minor concentration in individuals after LFHCC diet as compared to individuals after LFHCCn-3 diet. Taking into account that both diets were based on the same fat composition, the mentioned changes can be associated to the supplement included, high oleic acid sunflower oil or long chain *n-3* PUFA. The LFHCCn-3 diet also promoted a change in the profile of sphingomyelins as compared to the two diets based on high-fat content. Thus, the SM(d22:0/20:3) decreased its concentration in individuals after the LFHCCn-3 diet, while this SM increased its concentration in serum after the HSFA and HMUFA diets. One other representative effect was observed by comparing individuals after the LFHCCn-3 and HSFA diets since an oxidized PC(18:2-OH/16:0) increased its serum concentration in HSFA individuals (1.48 fold change), while this oxidized phospholipid decreased its serum level (<−2 fold change) after the PUFA supplemented diet.

It is also worth mentioning that two essential fatty acids — docosahexaenoic acid and linoleic acid— reported concentration changes in three diets. Thus, docosahexaenoic acid was found at increased concentration in LFHCCn-3 individuals, which included an n-3 PUFA supplement, while that was decreased in HMUFA cases. On the other hand, linoleic acid was detected at lower concentration in those controlled by the LFHCC diet as compared to control individuals.

3.3. Polar metabolic effects associated to the intervention diets

By analogy with the non-polar fraction, the influence of each diet on the polar metabolites in serum was evaluated by comparing the PRE sampling time versus the POS sampling time. Supervised analysis by multivariate comparison of polar tentative metabolites based on PLS-DA allowed establishing metabolic differences in serum profiles analyzed in individuals after the four intervention diets. In this fraction, prediction capability values were above 94% for the training set for the four diets, which supported the discrimination effect observed in each PLS scores plot (Figure 5).

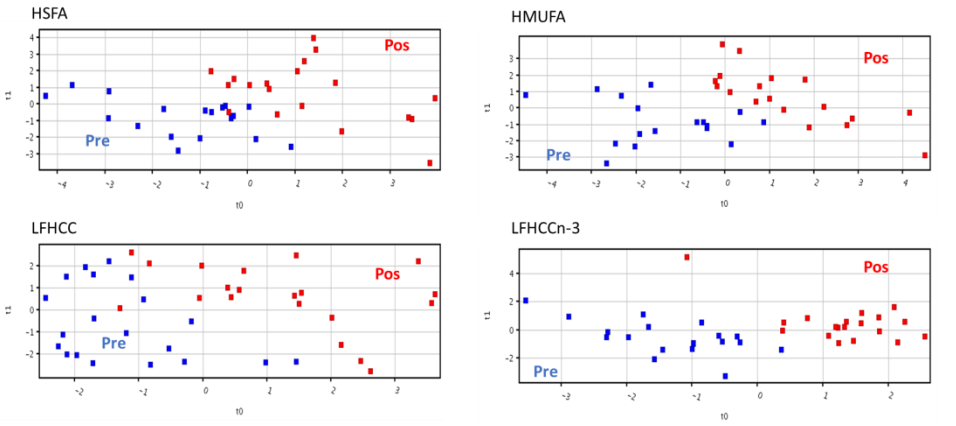


Figure 5. Scores graphs obtained after multivariate analysis by PLS-DA of the polar fraction for each diet, considering in each case PRE and POS times.

A fold change analysis was then executed to know the entities with capability to contribute to a higher extent to the separation between PRE and POS sampling times in each diet. Identification based on MS and MS/MS

information allowed elucidating the compounds that were altered in terms of relative concentration due to the intervention diet, as shows Table 2. The identified metabolites are classified by the fold change value, positive or negative, and diet. The significance level according to an ANOVA test is also shown. As can be seen, a common trend was observed in serum from all individuals subjected to the diets as compared to the pre-intervention state. Thus, octenoyl-L-carnitine resulted increased after the 12-weeks control period. Additionally, this effect was highly significant (99%) in the four groups of individuals. Common metabolites were found modified after the intake of one of either the low-fat diets or the high fat diets. Concretely, several amino acids, excepting ornithine, were only modified by intake of a high-fat diet, where the number of metabolites altered was higher (14–15) than in the case of the low fat diets (6–9). Apart from that, other relevant changes were observed for each diet. Thus, two critical differences were observed in the two high-fat content diets, HSFA and HMUFA. The first difference affects to levels of certain amino acids such as proline, creatine, betaine, valine and ornithine. These amino acids showed a different trend in the two groups of individuals. Ornithine was found at increased concentration in individuals after the HSFA diet, while those subjected to HMUFA diet decreased its concentration as compared to the pre-intervention state. The opposite effect was observed for the rest of amino acids, which means that valine, betaine and proline significantly increased their concentrations in HMUFA individuals, while proline, valine and creatinine decreased their concentration in HSFA diet. In addition to the phenomenon observed for ornithine, the same behavior was found for inosine, which increased its serum concentration in HSFA individuals (with statistical significance) while the HMUFA diet led to a relevant lowering effect (–1.82) in the concentration of this purine nucleoside. The mean area with the 95 percent of confidence levels for proline, methionine and valine for each time and diet are shown in Supplementary Figure 4.

Minor relevant changes were found in the polar fraction of serum from individuals subjected to the two low-fat content diets, LFHCC and LFHCCn-3. As previously emphasized, both diets resulted in increased concentration of

octenoyl-L-carnitine, but only in the LFHCC individuals were detected relevant effects for other two carnitine derivatives: hexanoyl-L-carnitine and acetylcarnitine,. Particularly, both carnitine analogues decreased their serum concentration due to the LFHCC diet. This diet also increased the concentration of ornithine as the HSFA diet.

Table 2. Polar metabolites that experienced a fold change above 1.25 by comparing physiological states before and after intervention diets.

	HSFA	**HMUFA**	**LFHCC**	**LFHCC(n-3)**
>2	Inosine* Paraxanthine L-ornithine Octenoyl-L-carnitine***	Octenoyl-L-carnitine***	Cholic acid	CMPF*** Octenoyl-L-carnitine***
1.75-2			Octenoyl-L-carnitine***	
1.5-1.75		Valine*		
1.25-1.5	LysoPC(16:0) CMPF Hippuric acid Bilirubin (1)	Betaine* Proline Paraxanthine CMPF Hippuric acid 2-Phenylacetamide	Ornithine	LysoPC(16:0)
(-1.25)-(-1.5)	Proline Creatinine Methionine* Glycocholic acid Bilirubin (2)	Aconitic acid Bilirubin (1) Methionine	Hexanoyl-L-carnitine* Acetylcarnitine*** Bilirubin (2)	Aconitic acid Inosine Hippuric acid Caffeine Hypoxanthine*
(-1.5)-(-1.75)	Caffeine Valine	Caffeine		Bilirubin (2)
(-1.75)-(-2)		Ornithine Inosine		
<(-2)				

***p-value < 0 .01
** 0.01<p-value<0.025
* 0.025<p-value
(1) and (2) bilirubin corresponds to bilirubin and bilirubin IXa, two indistinguishable isomers by mass spectrometry.

Two bile acids were also in the list of metabolites altered in terms of concentration: cholic acid and glycocholic acids, which are connected by a pathway regulated by choloylglycine hydrolase. The concentration of cholic acid was considerably increased (2.02) in individuals subjected to LFHCC, while glycocholic acid was decreased (–1.33) in HSFA diet. Finally, the LFHCCn-3 class was characterized by a decrease in the concentration of inosine, hippuric acid and

hypoxanthine (with a statistical significant decrease in the last case). On the other hand, the levels of some metabolites were increased or decreased only after intake of some of the diets under study. As an example, methionine decreased by diets with high-fat content (HSFA and HMUFA).

3.4. Pathways analysis

After identification of polar and non-polar metabolites that experienced changes in concentration due to the 12-weeks intervention period, an interpretation of the obtained results was carried out attending to the connection of biochemical pathways. Despite concentration changes of polar metabolites such as amino acids or carnitines were observed in relative terms, most alterations occurred on the metabolism of lipids, which was identified as the main variability source. Thus, the main metabolic impact caused by the intervention diets is reflected in the metabolism of glycerophospholipids and sphingolipids, which are the main components of cellular membranes, but also of blood lipoproteins, bile and lung surfactant. The influence of fatty acids intake on the qualitative and quantitative behavior of phospholipids had previously been reported [18]. In fact, concerning the three most abundant families of lipids present in serum, changes of the fatty acids composition in the diet have affected to phospholipids in a higher extent than to triglycerides and cholesteryl esters. The phenomenon seems to be logical since phospholipids constitute a family of compounds necessary for the absorption, transport and storage of lipids. With these premises, serum phospholipids could be proposed as sensitive markers of the fatty acid composition of diets. Among the metabolic changes altering the glycerophospholipids profile, it is worth mentioning that the serum levels of three of the most important phospholipids —PC(16:0/18:2), PC(18:0/16:2) and PC(18:0/18:2— have been affected by the diet; therefore they were able to set differences in individuals after diet intake. Additionally, the relative changes occurring in the profile of glycerophospholipids were not dependent on the fat energetic contribution that characterizes the diets. Thus, HSFA and LFHCC diets promoted a representative increase in relative concentrations of glycerophosphatidylcholines with two substituents, while the opposite effect was observed in

individuals subjected to LFHCCn-3 diet, which also evidenced a decrease in the level of sphingomyelins. The supplement of long chain n-3 PUFAs seems to be critical to explain these differences in metabolism.

Glycerophospholipids are synthetized from acyl-CoA, a group of coenzymes involved in fatty acids metabolism. The presence of PUFAs as substituents of glycerophosphatidylcholines contributed to set one other metabolic difference between groups of individuals: increased levels in HMUFA and LFHCC individuals versus decreased levels in HSFA and LFHCCn-3 individuals. Also, two PUFAs such as docosahexaenoic acid and linoleic acid, which are implicated in the metabolism of essential fatty acids, experienced differences among individuals subjected to different diets. The connection between the metabolism of fatty acids and glycerophospholipids establishes a global effect of intervention diets on the metabolism of lipids.

Choline, one other metabolite involved in the metabolism of glycerol-phospholipids, also revealed a statistical significant change in concentration due to the HSFA diet, while in HMUFA diet, betaine, which is synthesized by choline oxidation through conversion to betaine aldehyde, was that experiencing a change in concentration with statistical significance. Betaine is an intermediate in the biosynthesis of essential and non-essential amino acids. Thus, betaine plays a crucial role in the synthesis of L-carnitine that is the precursor of carnitines, the alteration of which was detected as one of the most representative metabolic effects caused by diets. Carnitine is an infrequent amino acid but an essential factor in the metabolism of fatty acids. This compound metabolizes fatty acids through the formation of acylcarnitines that generate acetyl-CoA and carnitine, thus completing the cycle of lipid degradation. Acetylcarnitine levels have also shown to be signiticantly decreased after LFHCC intake.

One other amino acid that was altered as a result of dietetic interventions was valine, whose degradation leads to leucine and, finally, to acetyl-CoA, which regulates the biosynthesis of fatty acids. In relation to valine, a variation was observed in individuals subjected to HSFA and HMUFA diets, since the levels of

this amino acid were increased in HMUFA individuals, while HSFA individuals described the opposite result.

Other two amino acids implicated in a common pathway, proline and ornithine, reported different results for LFHCC and HSFA individuals, since ornithine was increased in LFHCC individuals and decreased in HSFA individuals. Ornithine is produced in the urea cycle through arginine, but also it is the substrate to produce proline by the activity of the ornithine cyclodeaminase. The urea cycle is also connected to the citrate cycle, which could explain the decreased concentration of aconitic acid in LFHCCn-3 individuals as no modification of the metabolites involved in the urea cycle was observed. Methionine presented the same behavior in the two high-fat content diets as its concentration was decreased after the intervention period. However, its variation was only statistically significant in the case of the HSFA diet. Methionine is an intermediate of transmethylation reactions, being the methyl acceptor for the catabolism of betaine.

Two bile acids (cholic and glycocholic) were also affected by some diets. Bile acids facilitate the absorption, transport and excretion of fats and sterols in the intestine and liver. Therefore, it seems to be relevant that the type of fat administered to individuals could have an important effect on the metabolism of bile acids. Two diets revealed changes in the levels of bile acids: LFHCC, which increased the concentration of cholic acid, and HSFA, which promoted a relative decrease of glycocholic acid, the precursor of cholic acid.

Regarding purine base derivatives, inosine and hypoxanthine levels resulted modified after the intake of some of the diets. Concretely, inosine levels resulted significantly increased after HSFA intake, and decreased unsignificantly after HMUFA or LFHCCn-3 intake. Inosine is directly related to hypoxanthine, whose levels resulted significantly decreased after LFHCCn-3 intake. Hypoxanthine has demonstrated to induce oxidative stress related to lipid oxidation. In fact, the only oxidized phospholipid identified in the non-polar fraction — PC(18:2-OH/16:0)— decreased its concentration in individuals after LFHCCn-3

intervention, the same diet by which hypoxanthine was significantly decreased. On the other hand, hypoxanthine levels were significantly increased in individuals after HSFA intervention, the diet based on high content of SFA and the same diet by which inosine and the oxidized phospholipid were significantly and slightly increased, respectively.

Other compound with concentration modified after intake of some diets was hippuric acid, involved in the metabolism of phenylalanine. The concentration of hippuric acid was slightly increased after intake of high-fat content diets —HSFA and HMUFA— and slightly decreased after LFHCCn-3 diet. Hippuric acid has been associated to the metabolism of phenolic compounds, thus it could be associated to oxidation inhibition activity. Additionally, bilirubin 1 and bilirubin 2 (see Table 2) were found to increase and decrease, respectively their concentrations after HSFA intervention; bilirubin 1 experienced a decrease by HMUFA intake, while in LFHCC and LFHCCn-3 diets the decrease affected to bilirubin 2. Both bilirubin analogues are as potent physiological antioxidants. A less well-known metabolite, 3-carboxy-4-methyl-5-propyl-2-furanpropionic acid (CMPF), experienced a slight concentration increase in individuals subjected to one of the two high-fat content diets and the LFHCCn-3 diet. The occurrence of this metabolite in blood is associated to the consumption of fish, vegetables and fruits, being related with the metabolism of phospholipids [9]. Nevertheless, its origin is not clear and an endogenous presence by metabolism of furan fatty acids cannot be excluded.

4. Conclusions

A global nutrimetabolomics study has been developed to compare metabolic effects caused by four intervention diets differentiated in fat quantity and quality. The serum samples obtained before and after diet intake were fractioned by a mixture methanol–chloroform, and the polar and non-polar fractions evaluated separately. The comparison of serum metabolite profiles

obtained by LC–QTOF in high resolution mode has shown to be a suitable strategy to identify the most influenced metabolites after intervention diets.

Multivariate statistical analysis by PLS-DA revealed a slight discrimination among metabolic profiles representing the four post-intervention states, being the low fat content diet supplemented with n-3 fatty acids that providing the most discriminant metabolic pattern. A clear discrimination between the pre-intervention and post-intervention times for each diet was also observed.

The evaluation of the non-polar fraction led to know that phospholipids are the major fraction altered in the non-polar fraction after the intervention diets. On the other hand, the evaluation of the polar fraction revealed some amino acids, bile acids and acylcarnitines as some of the most affected metabolites in serum. In fact, the high fat content diets presented more amino acids with levels modified after diets, while only ornithine was slightly increased in one of the low fat content diets.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2012-37428). F. Priego-Capote is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks the MICINN for an FPU scholarship (AP2009-0499). The Reina Sofía Hospital (Córdoba, Spain) is also grateful to donors of serum samples.

References

[1] M. Mamas, W. Dunn, L. Neyses, R. Goodacre, Archives of Toxicology 85 (2011) 5-17.

[2] S. Moco, J. Vervoort, R. J. Bino, R. C. H. De Vos, R. Bino, TrAC Trends in Analytical Chemistry 26 (2007) 855-866.

[3] M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-McIntyre, P. Begley, K. Carroll, D. Broadhurst, A. Tseng, N. Swainston, I. Spasic, R. Goodacre, D. B. Kell, Analyst 134 (2009) 1322-1332.

[4] N. Psychogios, D. D. Hau, J. Peng, A. C. 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, D. S. Wishart, PLoS ONE 6 (2011) e16957.

[5] K. J. Boudonck, M. W. Mitchell, L. Német, L. Keresztes, A. Nyska, D. Shinar, M. Rosenstock, Toxicologic Pathology 37 (2009) 280-292.

[6] 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, Pharmacogenomics 9 (2008) 383-397.

[7] G. A. Theodoridis, H. G. Gika, E. J. Want, I. D. Wilson, Analytica Chimica Acta 711 (2012) 7-16.

[8] J. L. Griffin, Current Opinion in Chemical Biology 10 (2006) 309-315.

[9] J. L. Griffin R. A. Kauppinen, Journal of Proteome Research 6 (2006) 498-505.

[10] G. a. Graça, I. F. Duarte, B. J. Goodfellow, I. M. Carreira, A. B. Couceiro, M. d. R. r. Domingues, M. Spraul, L.-H. Tseng, A. M. Gil, Analytical Chemistry 80 (2008) 6085-6092.

[11] D. Vuckovic, Analytical and Bioanalytical Chemistry 403 (2012) 1523-1548.

[12] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrometry Reviews 26 (2007) 51–78.

[13] A. Koulman, G. Lane, S. Harrison, D. Volmer, Analytical and Bioanalytical Chemistry 394 (2009) 663-670.

[14] W. B. Dunn, N. J. C. Bailey, H. E. Johnson, Analyst 130 (2005) 606-625.

[15] K. Sandra, A. d. S. Pereira, G. Vanhoenacker, F. David, P. Sandra, Journal of Chromatography A 1217 (2010) 4087-4099.

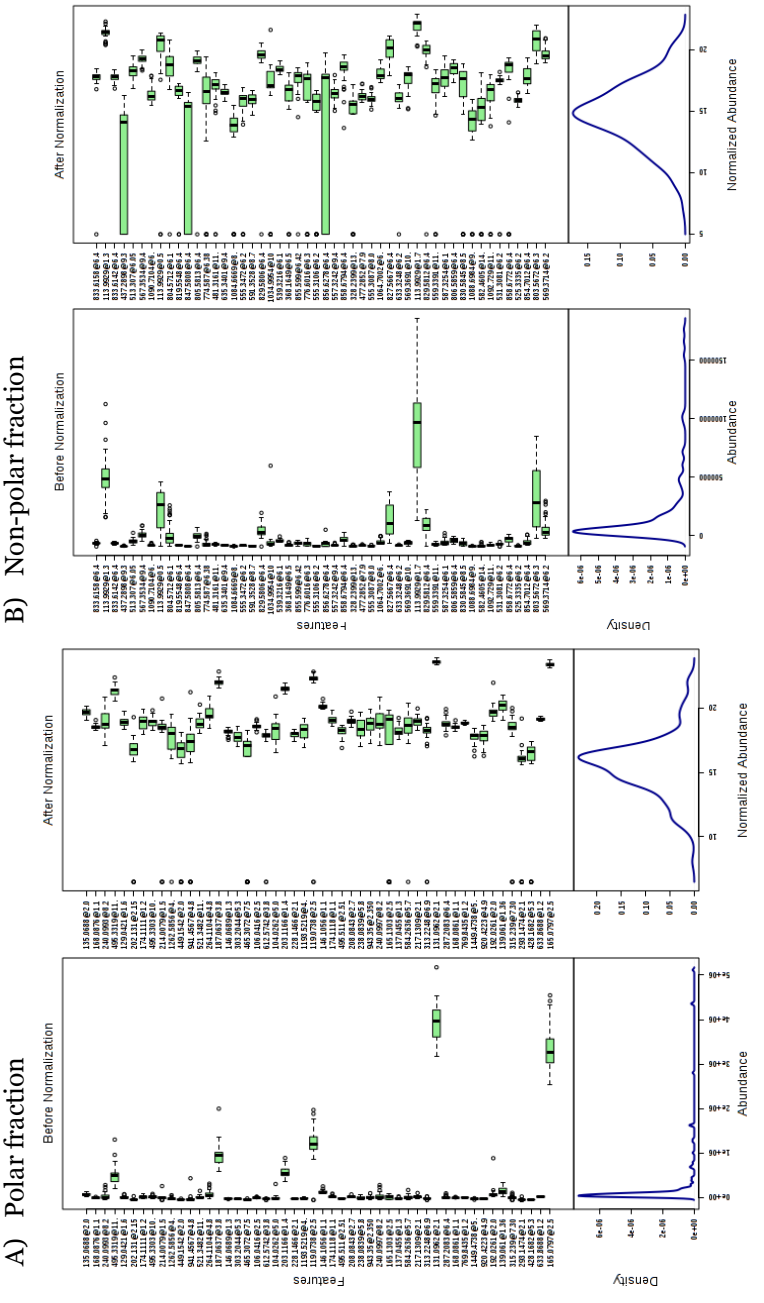
[16] D. I. Shaw, A. C. Tierney, S. McCarthy, J. Upritchard, S. Vermunt, H. L. Gulseth, C. A. Drevon, E. E. Blaak, W. H. M. Saris, B. Karlström, O. Helal, C. Defoort, R. Gallego, J. López-Miranda, D. Siedlecka, M. Malczewska-Malec, H. M. Roche, J. A. Lovegrove, British Journal of Nutrition 101 (2009) 750-759.

[17] P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini, P. Turano, Journal of Biomolecular NMR 49 (2011) 231-243.

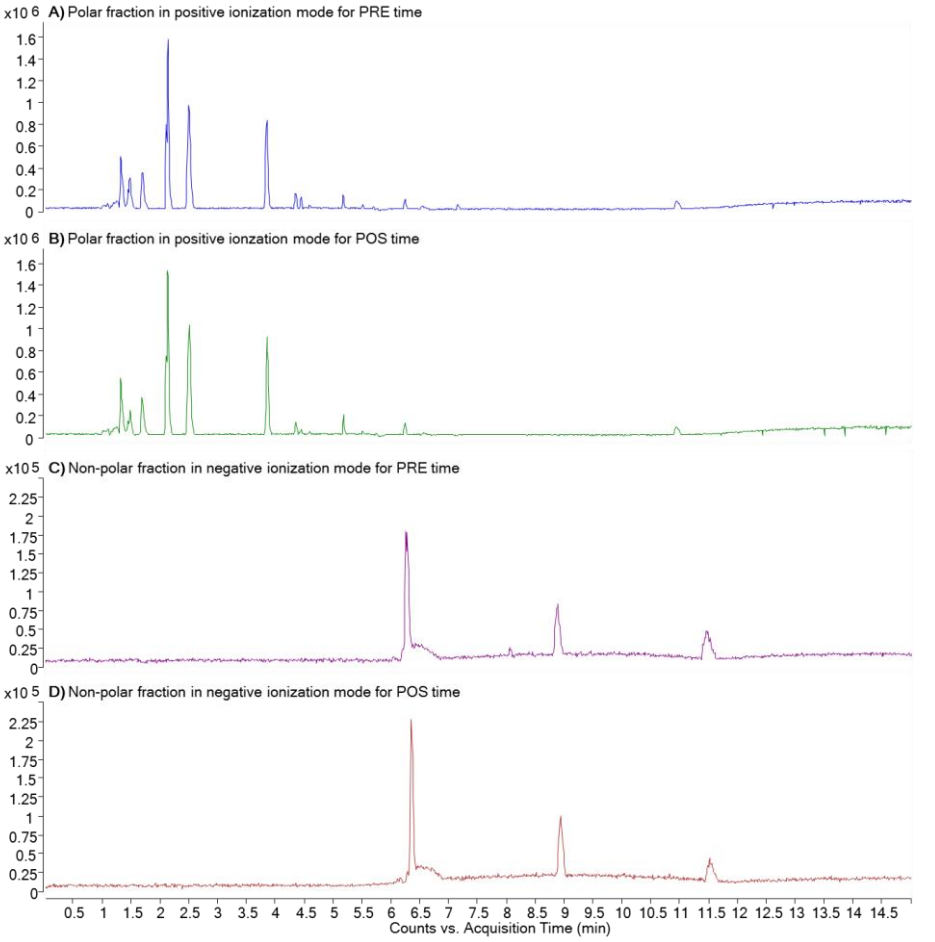
[18] P. Hlavatý, M. Kunesová, M. Gojová, E. Tvrzická, M. Vecka, P. Roubal, M. Hill, K. Hlavatá, P. Kalousková, V. Hainer, A. Zák, J. Drbohlav, Physiology Research 57 (2008) S57–S65.

[19] T. Niwa, Journal of Chromatography B: Analytical Technology and Biomedical Life Science 877 (2009) 2600–2606.

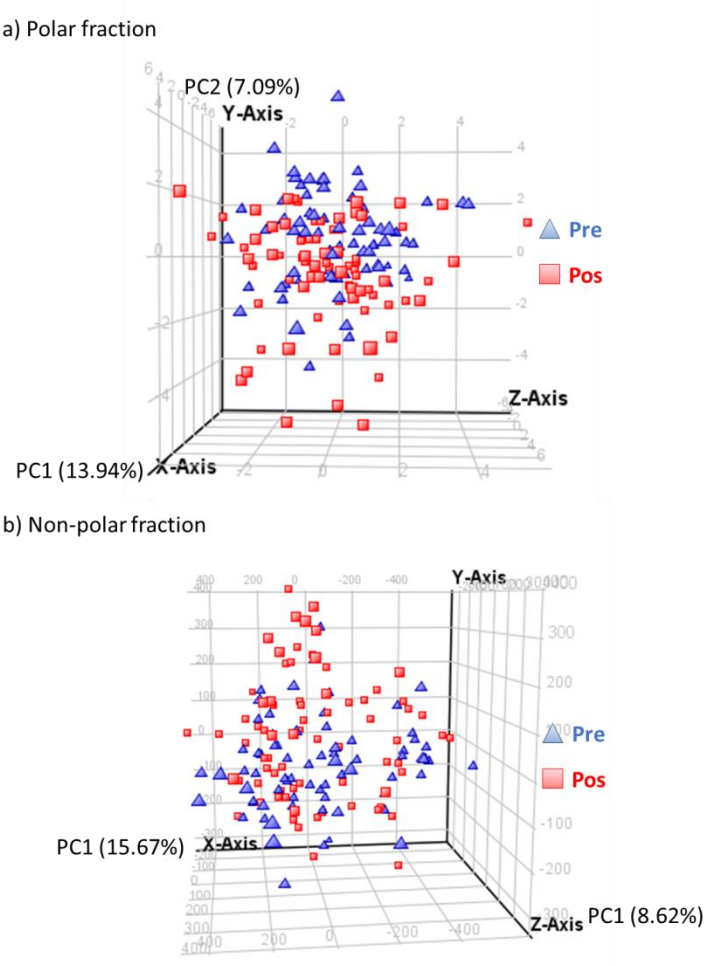
Supplementary material



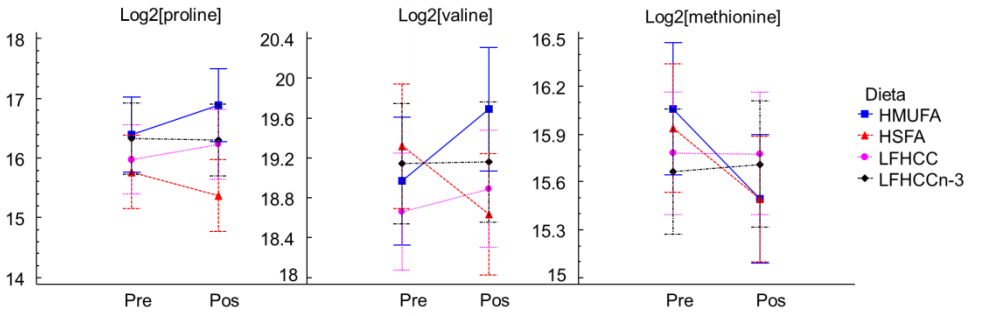
Supplementary Figure 1. Normalization effect in the peak area values of **(A)** polar and **(B)** non-polar metabolites. The box plots have been randomly selected from the complete dataset, while the density plots at the bottom have been built considering the profile of all the molecular features.



Supplementary Figure 2. Base peak chromatograms obtained by LC–MS/MS analysis of: **A)** the polar fraction of a pre-intervention sample in positive ionization mode; **B)** the polar fraction of a post-intervention sample in positive ionization mode; **C)** the non-polar fraction of a pre-intervention sample in negative ionization mode; and **D)** the non-polar fraction of a post-intervention sample in negative ionization mode.



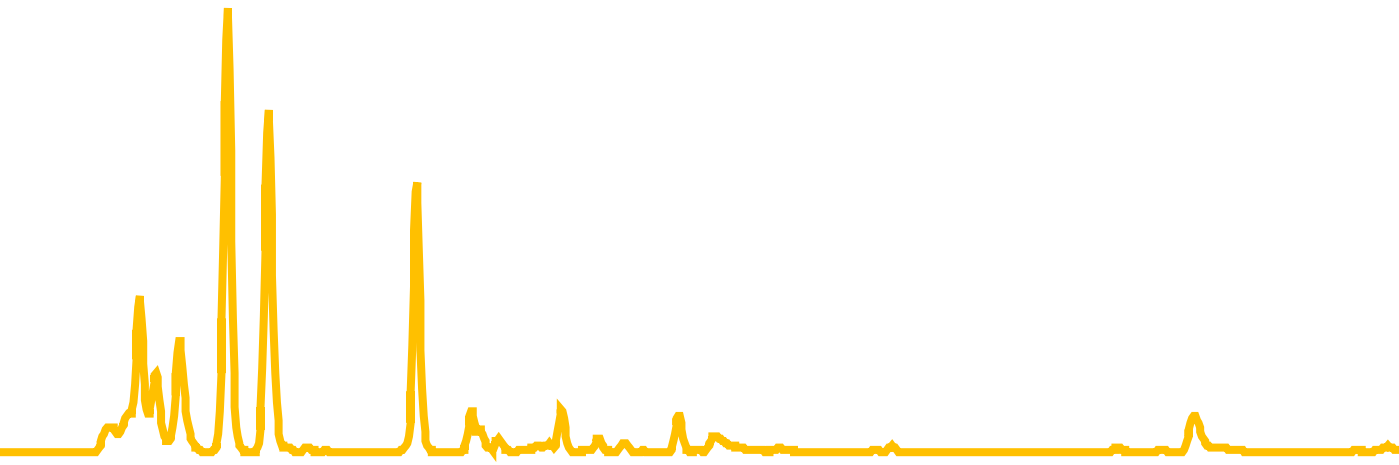
Supplementary Figure 3. Scores plot obtained for PRE and POS times considering all the diets after multivariate principal component analysis (PCA) for polar fraction (a) and non-polar fraction (b).



Supplementary Figure 4. Mean values and 95% confidence levels of the normalized area for proline, valine and methionine.

Chapter 5:

Metabolomic discrimination between patients with stable angina, non-ST elevation myocardial infarction, and acute myocardial infarct



Electrophoresis

34 (2013) 2827–2835

Metabolomic discrimination between patients with stable angina, non-ST elevation myocardial infarction, and acute myocardial infarct

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b], *J. G. Galache-Osuna*[c], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Department of Cardiology, Hemodynamic and Interventional Cardiology, University Hospital Miguel Servet, Zaragoza, Spain.*

Metabolomic discrimination between patients with stable angina, non-ST elevation myocardial infarction, and acute myocardial infarct

*M. Calderón-Santiago, F. Priego-Capote, J. G. Galache-Osuna, M. D. Luque de Castro**

Abstract

The ischemic cascade starts when atherosclerotic plaques decrease the supply of oxygen and substrates to cells and finalizes with myocardial infarction. These states have been here studied at metabolite level by optimization of a metabolomics profiling approach based on high-accuracy MS. For this purpose, serum samples from patients diagnosed with coronary artery disease and affected by stable angina or myocardial infarction (acute myocardial infarction or non-ST elevation myocardial infarction) were analyzed by LC–QTOF MS/MS after deproteinization to compare the profile of serum metabolites. The data set, composed by tentative molecular features detected in MS analyses, was filtered with statistical algorithms to remove entities resulting in redundant information. Tentative molecules were identified finding mainly lipids as statistically significant metabolites in the discrimination study due to their change in concentration. Lipids such as bile acid derivatives, phospholipids, and triglycerides were identified as relevant compounds for discrimination of individuals who suffered acute or non-ST elevation myocardial infarction from those suffering stable angina. The results achieved by this research could support the capability of metabolomics to go inside the study of artery diseases, and in addition to other omics disciplines could help to detect the occurrence of these disorders at initial stages or even to prognosticate their appearance.

1. Introduction

Nowadays, cardiovascular diseases (CVDs) are the leading cause of death worldwide with supremacy in low-and middle-income countries (80%) and with high significance in European countries (50%). Therefore, CVDs are a major cause of disability and contribute significantly to increase health care costs [1]. Cardiovascular diseases usually originate in atherosclerosis, a coronary artery disease characterized by the formation of lipid plaques in arteries restricting blood circulation to the heart. Atherosclerosis is the main cause of angina pectoris, a perceived symptom resulting from a mismatch of myocardial supply and demand caused by the partial obstruction of the artery. When the anaerobic threshold is crossed, the patient develops symptomatic angina pectoris (chest pain) which can appear after a physical effort or activity (angina stable) [2]. If this symptom is perceived with minimum activity and higher frequency, the suited term is unstable angina.

A rupture in the atherosclerotic plaque can generate a thrombus which can occlude normal blood flow. Complete thrombotic occlusion of a coronary artery causes myocardial necrosis and is termed acute ST-elevation myocardial infarction also known as acute myocardial infarction (AMI), whereas incomplete occlusion does not cause myocardial necrosis and is termed non-ST elevation myocardial infarction (NSTEMI). It is worth mentioning that unstable angina and NSTEMI are often considered together because they can be indistinguishable upon first appearance [3].

One of the main problems in clinical practice is the diagnosis of cardiovascular dysfunctions, which are commonly detected in a relatively advanced stage of the disease course [4]. The initial diagnostic approach for coronary artery disease encompasses a detailed patient history, a complete physical examination and an electrocardiogram. After this preliminary evaluation, laboratory blood tests, stress testing and invasive cardiac catheterization may be necessary to obtain further insight diagnostic and confirm the presence of atherosclerotic problems.

Different studies on atherosclerosis have been developed in the "omics" scene mainly aimed at the search for biomarkers and elucidation of the mechanisms through which myocardial infarction occurs. In fact, different proteomic databases of cardiac proteins have so far been constructed and alterations of several cardiac proteins in human myopathies have been described [5–7]. In the genomic area, research has been mainly focused on looking for genetic factors that may be associated with this disease susceptibility [8]. On the other hand, transcriptomics has also been used to identify specific patterns on myocardial tissue with different cardiovascular diseases (*e.g*. heart failure or cardiac transplants) [9, 10]. The latest developed omics, metabolomics, has contributed in a lesser extent to the study of atherosclerosis despite the metabolic changes that occur as a consequence of ischemia (*e.g*. accumulation of glucose, increased synthesis of prostacyclin and thromboxane, production of nitric oxide) [11]. Despite these well-known evidences, the metabolic impact associated to myocardial infarction is still unknown.

To date, there have been few metabolomics studies devoted to cardiovascular diseases. One of the first studies to uncover the potential of unbiased metabolomic profiles in predicting atherosclerosis patients was based on NMR spectroscopy analysis of blood samples [12]. This study included patients with unstable atherosclerosis or prior myocardial infarction. The model resulted in a 90% predictive capability for atherosclerosis patients on a major lipid region of the NMR spectra. However, a follow-up study was unable to keep this prediction results due to variability sources supported on gender and the use of3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors between the two populations under study [13]. These results emphasize the clinical variability of patient cohorts.

Other study utilized GC–MS to evaluate major changes in plasma metabolites between 9 patients with NSTEMI, 10 patients with stable atherosclerosis and 10 patients without coronary lesions [14]. Citric acid, 4-hydroxyproline, aspartic acid, and fructose levels were decreased, whereas lactate, urea, glucose, and valine levels were increased in atherosclerosis patients

as compared with the control group. Thus, metabolic changes were limited to polar metabolites.

Presently, no many biomarkers are used in the diagnosis of cardiovascular diseases —*e.g*. creatinin phosphokinase (CPK) or troponin for acute myocardial infarction. There are an ischemic cascade with elevation of specific markers at different stages of cardiovascular diseases, from plaque formation to necrosis and left ventricle remodeling, but most of the markers proposed so far are proteins, lipids or factors (*e.g*. placental growth factor, tissue factor) [15]. Nowadays, it is interesting to use metabolomics as a discrimination tool for disease diagnosis, given that many of the analytical tools used in metabolomics are automatable and on the whole high throughput, they lend themselves to screening populations of individuals for disease. There are still work to be done to fully elucidate the metabolic processes and changes occurring as a consequence of atherosclerosis. The objective of the present research was to find overall metabolic changes which could be used to differentiate three cardiovascular pathologies such as stable angina, NSTEMI and AMI by LC–QTOF/MS. This instrumental platform provides great advantages in metabolomics studies. The use of LC–MS/MS for identification and confirmatory analysis was also planned.

2. Materials and methods

2.1. Reagents

LC-grade methanol and MS-grade formic acid and acetonitrile (ACN) were purchased from Scharlab (Barcelona, Spain). Deionized water (18 mΩ • cm) from a Millipore Milli-Q water purification system was used.

2.2. Instruments and apparatus

A Sorvall Legend Micro 21R centrifuge (Thermo Scientific, Waltham, MA, USA) was used to centrifuge samples after deproteinization.

An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF tandem mass spectrometer equipped with a Jetstream® ESI source (Santa Clara, CA, USA) was used. The chromatographic eluting solution was monitored in high resolution mode.

2.3. Cohort selected for the study

A total of 189 patients affected by significant coronary artery lesion as atherosclerosis were included in the study. The main characteristics of the patients are shown in Supplementary table 1. As can be seen, the selected cohort was composed of 73 and 74 individuals diagnosed with stable angina and NSTEMI, respectively, while 42 of them had suffered previously an acute myocardial infarct. The average age of the cohort was 67±14, 79.0% of them male individuals, 51.6% smokers, 25.8% diabetic, 37.8% obese, 61.4% hypertense and 52.7% with hypercholesterolemia. Despite the variability of these factors, their proportion in the three considered sub-groups was quite similar. In fact, independence of clinical factors and cardiovascular pathologies was proved according to the Pearson's Chi Squared test with a 95% of confidence level. Atherosclerosis was diagnosed to all patients after evaluation through cardiac catheterization (angiographic stenosis revealed a reduction of the arterial lumen ≥ 70%). Blood extraction immediately preceded catheterization. Most patients were taken medication as follows: β-blocking agents (78%), statins (92%), antiplatelet agents (88%) and aspirin (98%).

2.4. Blood extraction and serum isolation

All steps from blood extraction to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004), which were supervised by specialized personnel from Miguel Servet Hospital (Zaragoza, Spain) that approved the experiments. Informed consent was obtained from all participants prior to sample collection.

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated for 30 min at room temperature to allow coagulation. Then, the tubes

were centrifuged at 2000 *g* for 15 min at 4°C to isolate the serum fraction (which was put in an ice bath prior to freezing within 2 h after collection). Serum was placed in plastic ware tubes and stored at −80 °C until analysis.

2.5. Sample treatment

Serum samples (100 μL) immersed in an ice bath were treated for deproteinization with methanol (200 μL), the most common solvent used for metabolomics profiling without sample fractionation. The mixture was shaken for 1 min and the precipitate removed after centrifugation for 5 min at 4 °C and 13800 *g*. The upper phase was collected in a vial that was placed in the LC autosampler for subsequent analysis.

2.6. LC–QTOF MS/MS analysis

Chromatographic separation was performed using a Mediterranea reversed phase C18 analytical column (100 mm x 0.46 mm i.d., 3 μm particle size) from Teknokroma (Barcelona, Spain) kept at 25 °C. Mobile phases were water (phase A) and ACN (phase B), both with 0.1% formic acid as ionization agent. The LC pump was programmed with a flow rate of 0.8 mL/min and the following gradient elution was carried out: 3% phase B was kept constant from min 0 to 2; from 3 to 100% of phase B from min 2 to 27; finally, 100% of phase B was kept from min 27 to 33.

The injection volume was 10 μL and the injector needle was rinsed with 70% methanol for 10 times between injections. Furthermore, the needle seat back was flushed for 15 s at a flow rate of 4 mL/min with 70% methanol to avoid cross contamination.

Jetstream ESI source parameters, operating in negative and positive ionization mode, were as follows: nozzle, capillary and fragmentor voltage were set at +/− 2 kV, +/− 3.5 kV and 175 V, respectively; N_2 nebulizer gas was flowed at 40 psi; N_2 drying gas flow rate and temperature were 8 L/min and 325°C; N_2 sheath gas flow rate and temperature were 11 L/min and 350°C. The instrument was calibrated and tuned according to procedures recommended by the

manufacturer. The data were collected in both centroid and profile modes at a rate of 2 spectrum per second in the extended dynamic range mode (2GHz). Firstly, the samples were analyzed in full scan MS mode in the *m/z* range 60–1100 for metabolite profiling. The reported data set was statistically processed to detect molecular features with a high significance to explain the observed variability. Once the most significant entities were recognized, the samples were analyzed in targeted MS/MS mode (*m/z* range 60–1100) to obtain structural information with identification purposes. To assure the desired mass accuracy of recorded ions, continuous internal calibration was performed during analyses with the use of signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated hexakis(1H, 1H, 3H-tetrafluoropropoxy)phosphazine or HP-921] in positive ion mode; in negative ion mode ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.000725 (formate adduct of HP-921) were used. Resolution provided by the instrument ranged from 15000 FWHM (Full Width at Half Maximum) at low masses (purine) and 30000 FWHM at high mass values (HP-921).

Samples were randomized injected twice to obtain chromatograms in both ionization modes. A blank sample was injected every 10 samples and the ESI source was cleaned every day to avoid carry over. A stability study was planned by repetitive injections of samples to avoid variability associated to disturbances in the metabolic profile. As can be seen in Supplementary figure 1, variability was found from the 13th injection, which sets the maximum period that samples can be at the autosampler without affecting their stability.

2.7. Data processing and statistical analysis

Raw data files were converted to mzData files using MassHunter Workstation software (version 3.01 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) and then the data from each polarity were separately processed in R statistical language (version 2.15.0, http://www.r-project.org/) using the open-free XCMS R-package (version 1.24, http://metlin.scripps.edu/xcms/index.php) [16,17] and the also open-free CAMERA R-package (version

1.12.0, http://www.bioconductor.org/packages/2.10/bioc/html/CAMERA.html). First, the XCMS package was used for processing mzData files of each ionization mode by extraction of potential molecular features (MFs) considering only ions exceeding 1000 counts peak height with a peak width between 10 and 60 s, a signal to noise threshold of 10 and 5 ppm of error in mass accuracy. Then, a non-linear alignment was executed using the "retcor" function with a degree of smoothing for local polynomial regression fitting of 0.5 and a symmetric fitting, process that used a re-descending M estimator with Tukey's biweight function. Then, the CAMERA package was used to correlate potential adducts (in positive ionization mode $[M+Na]^+$ and $[M+H]^+$;in negative ionization mode $[M+HCOO]^-$, $[M+Cl]^-$ and $[M-H]^-$) and isotopic peaks of the same molecular entity. Adducts formed by dehydration neutral loss were also considered. Thus, ions with identical elution profiles and related *m/z* values (representing different adducts or isotopes of the same compound) were grouped as a unique feature to remove redundant information. This process resulted in a data set containing the intensity values in the apex of chromatographic peaks for all molecular entities characterized by accurate mass and retention time (RT). Background contribution was removed by subtraction of MFs linked to plasticizers, solvent impurities and other contaminating compounds after analysis of blank sample (methanol) under identical instrument operation conditions.

The data set from each polarity was then imported to the Mass Profiler Professional (MPP) software package (version 2.0, Agilent Technologies, Santa Clara, CA, USA) for further processing. Data pretreatment was based on baselining to median across samples to remove background noise and normalization by logarithmic transformation to reduce relatively large differences among MF abundances. A fold-change algorithm was applied as filter to retain those molecular entities that experienced a change in relative concentration of 2.0 between pairs of classes. MPP software also allowed unsupervised and supervised analysis by PCA and PLS-DA of the data. In the case of PCA, data scaling by mean centering was used as pretreatment. Auto-scaling was selected in the case of PLS-DA. The validation model selected was N-Fold. With this model

the classes in the input data are randomly divided into N equal parts; N-1 parts are used for training, and the remaining one part is used for testing. The process repeats N times, with a different part being used for testing in an iterative process. 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. For all validations 10 repetitions and a fold number of 3 were selected.

Finally, an ANOVA test was executed for each metabolite to know the discrimination capability of each one among the three groups of patients considering significant a *p*-value lower than 0.05.

3. Results and discussion

3.1. Optimization of the LC–QTOF MS/MS analysis

Profiling analysis of serum metabolome from cardiovascular patients was carried out with a global approach based on non-targeted analysis. A simple deproteinization step with methanol was selected in this study to avoid sample fractionation which would increase the number of analysis. Chromatographic resolution was optimized with three different elution gradients, 22, 33 and 55 min, from aqueous solution (phase A) to acetonitrile (phase B) in both ionization modes. For the three methods tested, a 3% of phase B was kept for 2 min and, then, a gradient to 100% of B was programmed for 14, 25 and 47 min with a final step for 6 min of 100% B. Figure 1 includes base peak chromatograms (BPCs) provided by analysis of three aliquots of serum pool with the three chromatographic methods using negative ionization mode. According to the number of features extracted (3 replicates), the longest chromatographic method did not improve chromatographic resolution in any ionization modes. Therefore, theintermediate gradient method was adopted for analysis of samples from the complete cohort of individuals.

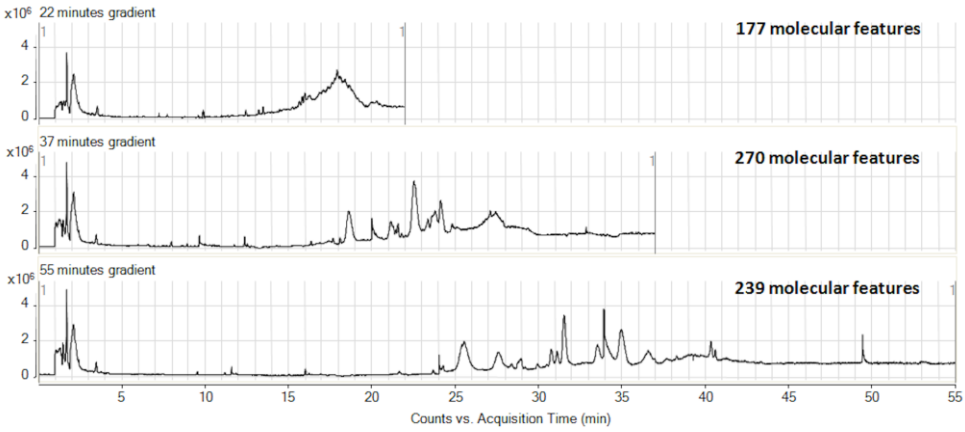


Figure 1. Base peak chromatograms obtained by the three chromatographic methods tested in negative ionization mode from three aliquots of serum pool. The number of molecular features obtained by each chromatogram are also included.

3.2. Data mining and pre-treatment

The alignment of molecular features is a crucial step prior to statistical analysis. The tolerance window for alignment employed was 0.2 min and 5 ppm for elution time and mass accuracy, respectively. Supplementary figure 2.A illustrates the curves for retention time deviation with elution time for all molecular entities after alignment. As can be seen chromatographic profiles were well-aligned demonstrating that chromatographic precision was quite good. One outlier was detected according to retention time deviation and, for this reason, this entity was removed. After data pretreatment, the data set obtained in the negative ionization mode contained 289 molecular entities. Supplementary figure 2.B shows the frequency histogram as a bars diagram plotting the number of entities with a certain frequency. The bars diagram reveals that most entities (92.7%) were detected in at least 90% of samples considered in this study, which supports a data set representative of the selected cohort. The data set obtained in the positive ionization mode did not reported significant differences among the groups under study, so these data were removed.

3.3. Statistical analysis by unsupervised and supervised methods

As mentioned above, myocardial infarction or heart attack implies myocardial cell death as a consequence of prolonged ischemia (impairment of blood flow with inadequate oxygen delivery to the heart muscle). Some heart attacks strike suddenly but it is also frequent to experience symptoms (angina) some hours, days or weeks in advance.

Despite some metabolic changes occurring in a myocardial infarction have been described [15,18], it may be of interest to compare metabolic profiles of different atherosclerosis stages to elucidate metabolic pathways of the process. For this purpose, a fold-change filter was applied to simplify the data set and eliminate redundant information. The filter removed those molecular features that did not experience a relative change in concentration of 2.0 between pairs of groups. The fold-change algorithm reduced the data set from 289 to 13 molecular features, which should explain a relevant part of the variability associated to the three groups. With this new data set, a principal component analysis was executed but no discrimination of the three groups was observed (Supplementary figure 3) due to overlapping of patients pertaining to the three groups. For this reason, PCAs were split into three studies: stable angina *versus* NSTEMI, stable angina *versus* AMI and NSTEMI *versus* AMI. Figure 2 illustrates the three-dimensional scores plot for the three independent studies. As can be seen, the two PCAs involving individuals diagnosed with stable angina clearly differentiated the two groups considered (Figure 2.A and 2.B). Therefore, relevant metabolic changes allowed differentiating patients diagnosed with the less severe diagnostic in atherosclerosis (stable angina) *versus* those diagnosed with NSTEMI and those who suffered an AMI episode. On the other hand, no discrimination was observed between individuals with NSTEMI and after an AMI episode (Figure 2.C). With these premises, prediction models based on PLS-DA were developed for the three studies to estimate the discrimination capability supported on the clinical diagnostic. Figure 3 shows the three-dimensional plots for each PLS model with the first three latent variables. As can be seen, PLS-DA plots confirmed the results obtained by PCA since clear discrimination was

observed in those cases involving individuals diagnosed with stable angina but not for the NSTEMI/AMI study. Prediction capability values expressed in percentage for training and validation sets of each PLS-DA are shown in Table 1.

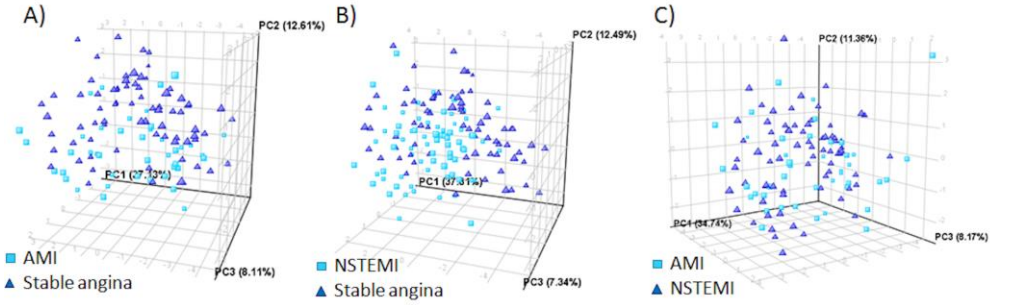


Figure 2. PCA scores plot obtained for AMI *versus* stable angina (**A**), stable angina *versus* NSTEMI (**B**), and AMI *versus* NSTEMI (**C**).

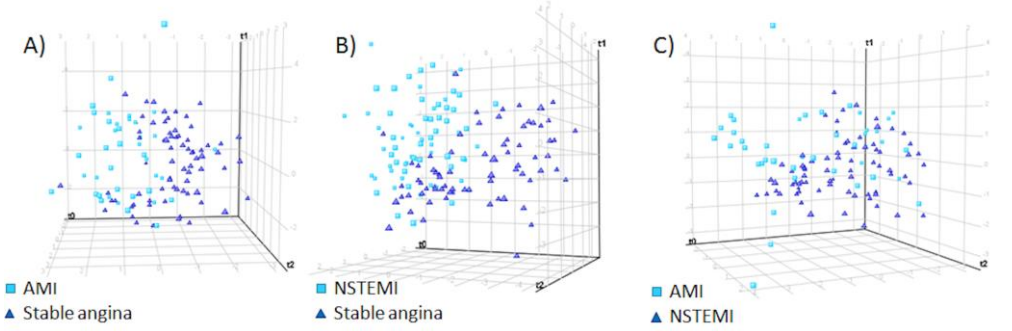


Figure 3. PLS-DA scores plot of the three models: AMI *versus* stable angina (**A**), stable angina *versus* NSTEMI (**B**), and AMI *versus* NSTEMI (**C**).

Therefore, supervised analysis by PLS-DA confirmed the capability of the panel of molecular features obtained after simplification of the data set to discriminate patients diagnosed with stable angina *versus* those who suffered more severe diagnostics: NSTEMI or AMI.

Table 1. Prediction capability for the three PLS-DA models considered.

	Model 1		**Model 2**		**Model 3**	
	AMI	**Stable angina**	**NSTEMI**	**Stable angina**	**AMI**	**NSTEMI**
Model validation						
Sensitivity (%)	76.2	72.6	74.0	79.7	56.7	59.5
Specificity (%)	72.6	76.2	79.7	74.0	59.5	56.7
Recognition ability (%)	75.68		76.9		57.7	
Model Training						
Sensitivity (%)	74.0	71.4	78.0	83.8	66.2	69.0
Specificity (%)	71.4	74.0	83.8	78.0	69.0	66.2
Recognition ability (%)	73.0		80.9		67.2	

3.4. Identification of significant metabolites contributing to explain diagnostics

Targeted analysis of significant molecular features was carried out by tandem mass spectrometry with the QTOF spectrometer for identification of metabolites, which would allow interpreting the connection between metabolic changes and clinical diagnostic. The HMDB and the METLIN databases were used to support the identification step. The identified metabolites are listed in Table 2 with analytical information such as retention time, *m/z* value of the detected adduct, error expressed in ppm and the main fragment ions used for identification. According to the fold-change algorithm, cholic acid was the most significant metabolite to discriminate patients affected by AMI from those diagnosed with stable angina. On the other hand, lysophosphocholine (16:0) was the metabolite with the highest relative change in terms of concentration for discrimination of patients diagnosed with NSTEMI from those with angina stable.

Cholic acid is one of the major bile acids biosynthesized from cholesterol and transformed into other secondary bile acids, one of them being deoxycholic acid [19]. Bile acids are usually the form in which the excess of cholesterol in the body is eliminated, so low levels of bile acids are generally related to high levels of

Table 2. List of identified metabolites including retention time (RT), monoisotopic molecular weight (M_w), *m/z* value of the detected adducts and product ions, and the error expressed in ppm. The average intensity of all metabolites in each group under study, expressed as mean ± standard deviation (SD), is also included.

Compound name	M_w	RT *(min)*	Parent ions *(m/z)*	Product ions *(m/z)*	Δppm	Intensity average *(Area±SD)*		
						AMI	NSTEMI	Stable angina
Phospholipids and lysophospholipids								
LysoPC (16:0)	495.3325	15.91	540.33133 $[M+FA-H]^-$	480.3086 255.2356 392.2388	1.11	$2.57{\cdot}10^6 \pm 2.48{\cdot}10^6$	$2.79{\cdot}10^6 \pm 2.83{\cdot}10^6$	$2.92{\cdot}10^6 \pm 2.82{\cdot}10^6$
PS (14:1/15:1) or PS (14:1/14:1) methyl ester	689.4268	16.37	734.42426 [M+FA-H]	167.0190	1.00	$4.23{\cdot}10^3 \pm 4.38{\cdot}10^3$	$4.73{\cdot}10^3 \pm 7.67{\cdot}10^3$	$5.65{\cdot}10^3 \pm 6.14{\cdot}10^3$
LysoPC(18:1)	521.3481	18.73	566.34593 $[M+FA-H]^-$	506.3183	0.65	$7.40{\cdot}10^4 \pm 1.43{\cdot}10^5$	$9.65{\cdot}10^4 \pm 1.27{\cdot}10^5$	$1.14{\cdot}10^5 \pm 1.70{\cdot}10^5$
PC(16:0/5:0(CHO))	593.3687	20.71	638.36699 $[M+FA-H]^-$	196.9373 339.1324 255.2324 490.2696 115.0395	0.14	$1.14{\cdot}10^4 \pm 1.33{\cdot}10^4$	$1.87{\cdot}10^4 \pm 1.98{\cdot}10^4$	$1.61{\cdot}10^4 \pm 2.43{\cdot}10^4$
LysoPC(18:2)	519.3325	24.30	554.30182 $[M+Cl]^-$ 504.30937 $[M-CH_3]^-$	279.2328 311.1139	1.24	$5.80{\cdot}10^4 \pm 5.74{\cdot}10^4$	$7.57{\cdot}10^4 \pm 6.34{\cdot}10^4$	$5.92{\cdot}10^4 \pm 5.74{\cdot}10^4$
Other lipids								
TG (18:3/20:4/18:3)	898.7050	16.27	448.30628 $[M-2H]^{-2}$	301.6223 150.1414 300.2095	1.29	$3.14{\cdot}10^3 \pm 5.33{\cdot}10^3$	$6.36{\cdot}10^3 \pm 1.15{\cdot}10^4$	$1.17{\cdot}10^4 \pm 2.91{\cdot}10^4$
PA (10:1/18:1)	588.3791	27.97	293.17925 $[M-2H]^{-2}$	96.9691 293.1721 295.2637	10.57	$1.31{\cdot}10^5 \pm 1.53{\cdot}10^5$	$1.44{\cdot}10^5 \pm 1.78{\cdot}10^5$	$6.26{\cdot}10^4 \pm 9.31{\cdot}10^4$
Bile acids								
Bilirubin	584.2635	31.66	583.25654 $[M-H]^-$	518.2171 263.1359 381.1509 134.1254	0.20	$2.17{\cdot}10^5 \pm 1.97{\cdot}10^5$	$2.37{\cdot}10^5 \pm 1.68{\cdot}10^5$	$1.39{\cdot}10^5 \pm 1.36{\cdot}10^5$
Cholic acid	408.2876	17.73	407.27885 $[M-H]^-$	233.2437	3.60	$1.97{\cdot}10^4 \pm 6.01{\cdot}10^4$	$2.07{\cdot}10^4 \pm 5.91{\cdot}10^4$	$2.20{\cdot}10^4 \pm 4.06{\cdot}10^4$
Deoxycholic acid	392.2927	20.95	391.28510 $[M-H]^-$	273.6710	0.78	$1.67{\cdot}10^4 \pm 2.87{\cdot}10^4$	$2.15{\cdot}10^4 \pm 3.26{\cdot}10^4$	$2.78{\cdot}10^4 \pm 3.52{\cdot}10^4$
Vitamin								
25-Hydroxyvitamin D3	400.6371	28.28	445.33232 $[M+FA-H]^-$	401.3450 383.3282 160.1523	0.02	$4.08{\cdot}10^4 \pm 4.20{\cdot}10^4$	$4.75{\cdot}10^4 \pm 4.33{\cdot}10^4$	$3.25{\cdot}10^4 \pm 4.00{\cdot}10^4$
Unknown metabolites								
$C_{26}H_{42}O_9$ or $C_{22}H_{45}NO_9P$	498.2829 498.2832	16.43	497.27492 $[M-H]^-$	453.1804 268.9535	1.42 2.04	$3.61{\cdot}10^3 \pm 4.45{\cdot}10^3$	$3.48{\cdot}10^3 \pm 4.38{\cdot}10^3$	$8.79{\cdot}10^3 \pm 1.38{\cdot}10^4$
$C_{18}H_{28}O_9S$	356.1657	29.46	355.15819 $[M-H]^-$	297.1496 311.1653	0.78	$7.11{\cdot}10^4 \pm 5.65{\cdot}10^4$	$7.73{\cdot}10^4 \pm 5.63{\cdot}10^4$	$4.45{\cdot}10^4 \pm 4.39{\cdot}10^4$

cholesterol, one of the CVD risk factors. The levels of cholic and deoxycholic acids were not significantly related with the presence of hypercholesterolemia (*p*-values 0.5986 and 0.1005, respectively). Nevertheless, a positive relationship between hypercholesterolemia and deoxycholic levels was observed as Supplementary figure 4 shows. Furthermore, it has been demonstrated that coronary artery disease patients have significantly decreased bile acid excretion levels as compared with normal population [20]. Supplementary figure 5 shows that patients with AMI presented the lowest level of both bile acids, while patients with stable angina presented the highest levels; however, a low discrimination capability is obtained by both metabolites.

One other identified bile metabolite was bilirubin which is the end-product of porphyrin rings degradation catalyzed by hemeoxygenase (HO). This enzyme generates free redox active iron, carbon monoxide and biliverdin, which is metabolized to bilirubin. One of the two HO isoforms, the HO-1, is a stress-responsive enzyme highly induced by many agents such as cytokines, endotoxin, heavy metals, nitric oxide and the own hemesubstrate [21]. An excess of nitric oxide released during occlusion of artery lumen could be responsible for the increase of bilirubin concentration in AMI and NSTEMI patients (Figure 4.A), which endows this metabolite with high discrimination capability (*p*-value 0.001). HO-1 represents a key defensive mechanism against further oxidative injury by virtue of the anti-inflammatory and antioxidant capacities of CO, biliverdin and bilirubin.

On the other hand, lysophosphocholines are produced by hydrolysis of lysophophatidylcholines, resulting from cholesterol by the action of cholesterol acyltransferases. LysoPC (16:0) is one of the most common lysoPCs formed by a chain of palmitic acid, which is present in animal fats, in vegetable and fish oils.

Apart from lysoPC(16:0), other phospholipids were identified in the resulting panel, which emphasizes the role of phospholipids in the development of the coronary artery disease. The metabolism of phospholipids has previously been related to cardiovascular disorders [22]. In fact, lysoPCs play a key role in

plaque inflammation and vulnerability, whereas oxidized phospholipids tend to accumulate at sites where oxidative stress is especially involved. Among them, it is worth mentioning areas with atherosclerotic lesions, hyperlipidemic plasma or with low concentration of high-density lipoproteins. Additionally, the reverse cholesterol transport is inhibited, thus contributing to the development of hypercholesterolemia and atherosclerosis [23]. Other relevant lysoPC identified in the panel of significant metabolites were lysoPC(18:2) and lysoPC(18:1) which, together with lysoPC(16:0), constitute three of the main phospholipids present in human plasma. Supplementary figure 6 illustrates the average levels obtained for the three lysoPCs in the three groups considered in this research. An oxidized phosphocoline —PC(16:0/5:0(CHO)— was also detected with a relative change in concentration in NSTEMI patients *versus* stable angina patients, supporting the importance of phospholipids oxidation in the atherosclerotic cascasde [24]. Supplementary figure 6.D describes the average levels of this oxidized phospholipid in the three groups included in the study.

Two additional phospholipids were identified as contributing metabolites: a phosphatidylserine (compound 1), and PA(10:1/18:1). Phosphatidylserines are located mainly on the inner monolayer surface of the plasma cell membrane and are the most abundant anionic phospholipids contributing to interfacial effects in membranes involving non-specific electrostatic interactions. The main cause of their presence in serum is connected to cellular apoptosis, which takes place after a myocardial infarction or during platelet activation [25]. In this case, two identification options were possible for *m/z* 734.4243: the $[M+FA-H]^-$ adduct of either methyl ester PS(14:1/14:1) or PS(14:1/15:1), the latter related with an exogenous factor such as diet. Thus, detection of C:15 fatty acid is based on the intake of dairy products and milk [26]. On the other hand, phosphatidic acids (PAs) are main components of cellular membranes and intermediates in the biosynthesis of triacylglycerols and phospholipids. PAs are involved in many processes: platelet aggregation, smooth muscle contraction, *in vivo* vasoactive effects, chemotaxis, expression of adhesion molecules, increased tight junction permeability of endothelial cells, induction of stress fibres,

modulation of cardiac contractility, and many others [27]. As can be seen in Figure 4.B and Supplementary figure 7.A, PA(10:1/18:1) and PS(compound 1) levels were lower in patients with stable angina than in those with AMI episode or NSTEMI diagnostic, but only PA(10:1/18:1) is endowed with high discrimination capability (*p*-value 0.002).

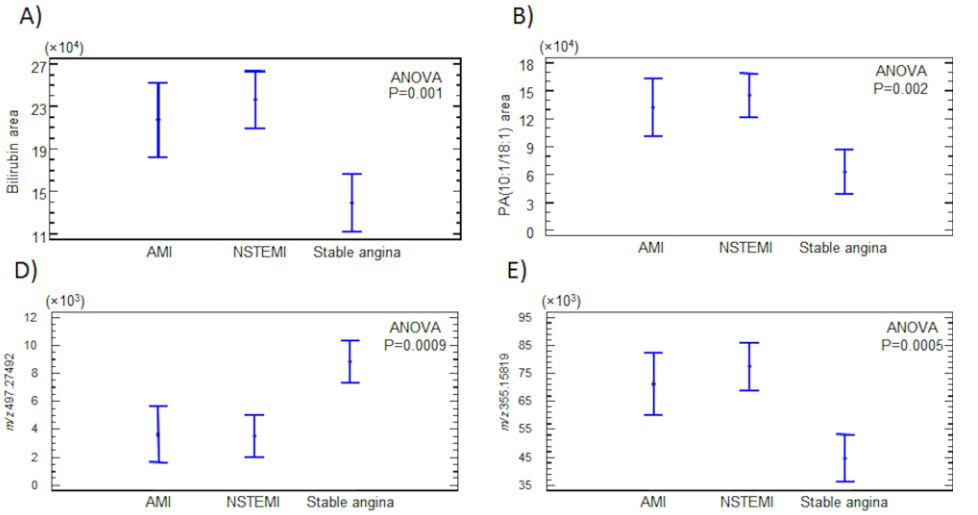


Figure 4. Means and 95% least significance difference (LSD) intervals for each group of patients for the metabolites: bilirubin (**A**), PA(10:1/18:1) (**B**), phosphatidylserine (**C**), *m/z* 497.27492 (**D**) and *m/z* 355.15819 (**E**). P-value of ANOVA test for those metabolites, and the cardiovascular disease as categorical factor are also shown.

The triglyceride TG(18:3/20:4/18:3) was also identified with a lower concentration in individuals who suffered an AMI episode or were diagnosed with NSTEMI as compared with individuals suffering from stable angina (see Supplementary figure 7.B). Triglyceride concentrations have been widely proposed as predictors of myocardial infarction as their levels seem to be related with the probability of cardiovascular events [28].

One other interesting lipid metabolite was 25-hydroxy metabolite of vitamin D3, which is the main circulating form of vitamin D3. Deficiency of this vitamin has been linked to an increased risk of hypertension, congestive heart failure, peripheral arterial disease, myocardial infarction, stroke, and related mortality, even after adjustment for traditional cardiovascular risk factors.

Clinical and epidemiological studies have evidenced that vitamin D may also be associated with several indices of vascular function, including the development and progression of atherosclerotic cardiovascular disease [29]. This fact could partially explain the significance of the monohydroxy metabolite of vitamin D3. Supplementary figure 8 shows the main levels of this metabolite in the three groups of patients.

Finally, two additional metabolites were not identified according to their MS/MS spectra. In this case, the objective was to calculate the molecular formula with the best isotopic distribution adjustment for the precursor ion and also for the most representative product ions. Thus, the precursor ion *m/z* 497.27492 eluted at 16.4 min was assigned to two potential molecular formulas ($C_{26}H_{42}O_9$ and $C_{22}H_{45}NO_9P$) with accuracy error of 1.42 and 2.04 ppm, respectively. Figure 4.E illustrates the high discrimination capability (*p-value* 0.0009) of this metabolite for the three classes under study. On the other hand, the precursor ion at *m/z* 355.15819 was eluted at 29.4 min and, for this reason, it should be a lipid metabolite. The tentative molecular formula for this metabolite was $C_{18}H_{28}O_5S$ with accuracy error of 0.78 ppm, which could fit to a sulfate derivative. The concentration levels for this metabolite in the three classes considered are shown in Figure 4.D, which explains its high discrimination capability (*p*-value 0.0005).

4. Conclusions

A group of 13 metabolites able to discriminate between different stages of atherosclerosis has been identified, most of them lipids and some bile acids as cholic or deoxycholic acids. Bilirubin has shown to be one of the most influential metabolites as it is able to discriminate between patients with stable angina and patients with AMI or NSTEMI with a *p*-value of 0.0012. Some of those compounds had previously been related to cardiovascular diseases or oxidative stress, which has shown to be very influential on atherosclerosis. It would be interesting to use other techniques as NMR or GC–MS to identify the unknown metabolites, thus assessing the results of this study.

Metabolomics has shown to be a very useful strategy to study metabolic changes relevant in cardiovascular disease and to propose new biomarkers to discriminate between different stages of atherosclerosis.

Acknowledgements

The Spanish Ministerio de Educación y Ciencia (MICINN) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2009-07430). F. Priego-Capote and M. Calderón Santiago are also grateful to the MICINN for a Ramón y Cajal Contract (RYC-2009-03921) and an FPU scholarship (AP2009-0499), respectively. The Miguel Servet Hospital (Zaragoza, Spain) is also thanked for donors' serum samples.

References

[1] Petersen, S., Peto, V., Rayner, M., Leal, J., Luengo-Fernández, R., Gray, A., European Cardiovascular Disease Statistics, British Heart Foundation, London, 2005.

[2] Rimmerman, C.M., in: Carey, W.D. (Ed.), *Current Clinical Medicine*, Saunders, Philadelphia 2010.

[3] Antman, E.M., Cohen, M., Bernink, P.J.L.M., McCabe, C.H., Horacek, T., Papuchis, G., Mautner, B., Corbalan, R., Radley, D., Braunwald, E., *JAMA-J. Am. Med. Assoc.* 2000, *284(7)*, 835–842.

[4] Barderas, M.G., Laborde, C.M., Posada, M., Cuesta, F., Zubiri, I., Vivanco, F., Alvarez-Llamas, G., *J. Biomed. Biotechnol.* 2011, doi: 10.1155/2011/790132.

[5] O'Donoghue, M., de Lemos, J.A., Morrow, D.A., Murphy, S.A., Buros, J.L., Cannon, C.P., Sabatine, M.S., *Circulation* 2006, *114*, 550–557.

[6] McCord, J., Nowak, R.M., Hudson, M.P., McCullough, P.A., Tomlanovich, M.C., Jacobsen, G., Tokarski, G., Khoury, N., Weaver, WD., *Ann. Emerg. Med.* 2003, *42*, 343–350.

[7] Devereux, R.B., Alonso, D.R., Lutas, E.M., Gottlieb, G.J., Campo, E., Sachs, I., Reichek, N., *Am. J.Cardiol.* 1986, *57*, 450–458.

[8] Schaefer, A.S., Richter, G.M., Groessner-Schreiber, B., Noack, B., Nothnagel, M., Mokhtari, N.E., Loos, B.G., Jepsen, S., Schreiber, S.,*PLoS Genet*.2009,*5(2)*,e1000378.

[9] Rudiger, A., Dyson, A., Felsmann, K., Carré, J.E., Taylor, V., Hughes, S.,Clatworthy, I., Protti, A.,Pellerin, D.,Lemm, J., Claus, R.A., Bauer, M., Singer, M., *Clin. Sci.* 2012, doi: 10.1042/CS20120334.

[10] Heidecker, B., Lamirault, G., Kasper, E.K., Wittstein, I.S., Champion, H.C., Breton, E., Russell, S.D., Hall, J., Kittleson, M.M., Baughman, K.L., Hare1, J.M., *Eur. Heart J.* 2010, *31*, 1188–1196.

[11] Goonewardena, S.N., Prevette, L.E., Desai, A.A., *Curr. Atheroscler. Rep.* 2010, *12*, 267–272.

[12] Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E., Grainger, D.J., *Nat. Med.* 2002, *8*, 1439–1444.

[13] Kirschenlohr, H.L., Griffin, J.L., Clarke, S.C., Rhydwen, R., Grace, A.A., Schofield, P.M., Brindle, K.M., Metcalfe, J.C., *Nat. Med.* 2006, *12*, 705–710.

[14] Vallejo, M., García, A., Tuñón, J., García-Martínez, D., Angulo, S., Martin-Ventura, J.L., Blanco-Colio, L.M., Almeida, P., Egido, J., Barbas, C., *Anal. Bioanal. Chem.* 2009, *394*, 1517–1524.

[15] Ingelsson, E., Vasan, R.S., in: Graham, I.M., D'Agostino, R.B. (Eds.), *Therapeutic strategies in cardiovascular risk*, Clinical Publishing, Oxford, 2008, pp. 111–136.

[16] Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G., *Anal. Chem.* 2006, *78*, 779–787.

[17] Tautenhahn, R., Böttcher, C., Neumann, S., *BMC Bioinformatics* 2008, *9*, 504.

[18] Bling, R.J., *CMLS-Cell. Mol. Life Sci.* 2001, *58*, 351–355.

[19] Russell, D.W., Setchell, K.D.R., *Biochem.*1992, *31(20)*, 4737–4749.

[20] Charach, G., Grosskopt, I., Rabinovich, A., Shochat, M., Weintraub, M., Rabinovich, P., *Therap. Adv. Gastroenterol.* 2011, *4(2)*, 95–101.

[21] Otterbein, L.E., Choi, A.M.K., *AJP – Lung Physiol.* 2000, *279(6)*, L1029–L1037.

[22] Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., DuGar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.M., Wu, Y., Schauer, P., Smith, J.D., Allayee, H., Tang, W.H.W., DiDonato, J.A., Lusis, A.J., Hazen, S.L., *Nature* 2011, *472*, 57–63.

[23] Ashraf, M.Z., Kar, N.S., Chen, X., Choi, J., Salomon, R.G., Febbraio, M., Podrez, E.A., *J. Biol. Chem.* 2008, *283(16)*, 10408–10414.

[24] Tsimikas, S., Brilakis, E.S., Miller, E.R.m McConnell, J.P., Lennon, R.J., Kornman, K.S., Witztum, J.L., Berger, P.B., *N. Engl. J. Med.* 2005, *353*, 46–57.

[25] Bennett, M.R., *Heart* 2002, *87*, 480–487.

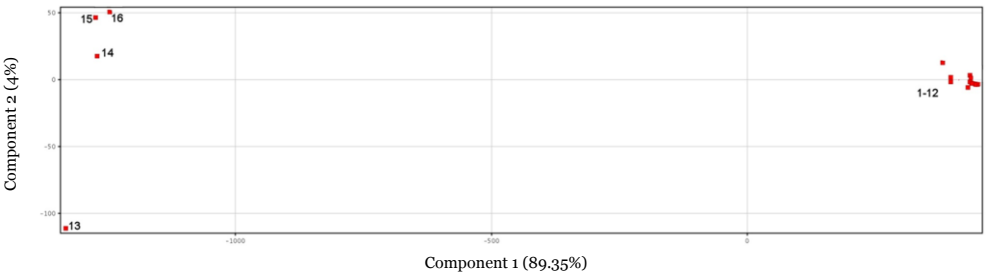
[26] Arab, L., *J. Nutr.* 2003, *133(3)*, 9255–9325.

[27] Park, K.A., Vasko, M.R., *Trends Pharmacol. Sci.* 2005, *26(11)*, 571–577.

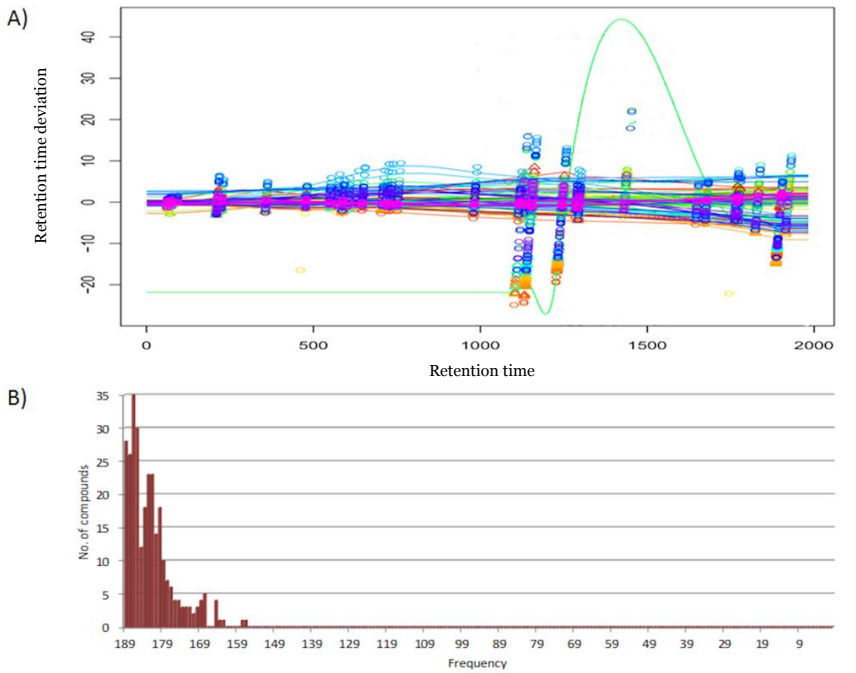
[28] Lindquist, P., Bengtsson, C., Lissner, L., Björkelund, C., *J. Intern. Med.* 2002, *251(6)*, 484–489.

[29] Brewer, L.C., Michos, E.D., Reis, J.P., *Curr. Drug Targets* 2011, *12(1)*, 54–60.

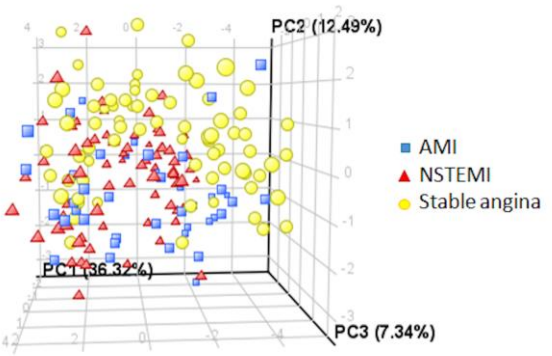
Supplementary material



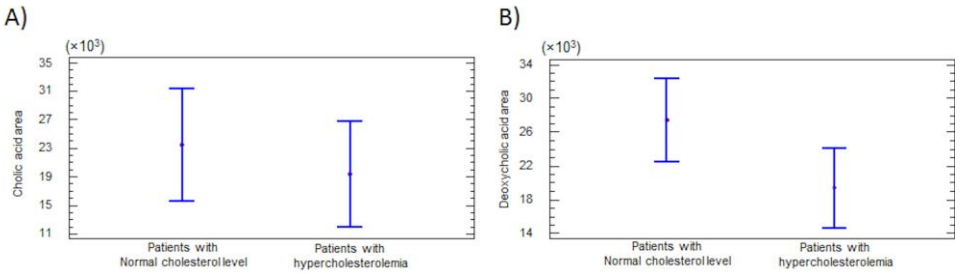
Supplementary figure 1. PCA comparing metabolic profiles associated to serum by repetitive injections of one aliquot in LC–MS in positive ionization mode.



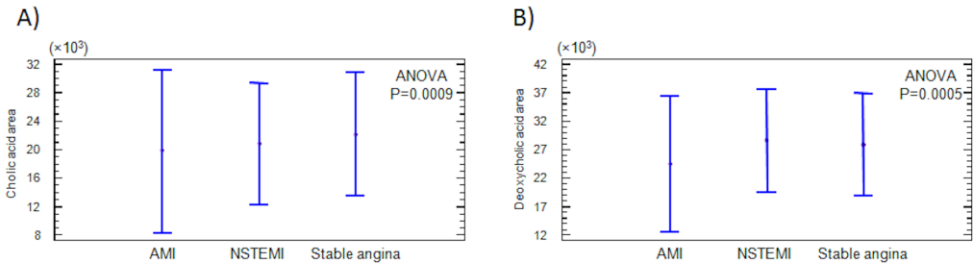
Supplementary figure 2. A. Retention time deviation curves obtained for all molecular entities after alignment as a function of the elution time. **B**. Bars diagram of frequency histogram plotting the number of entities with a certain frequency.



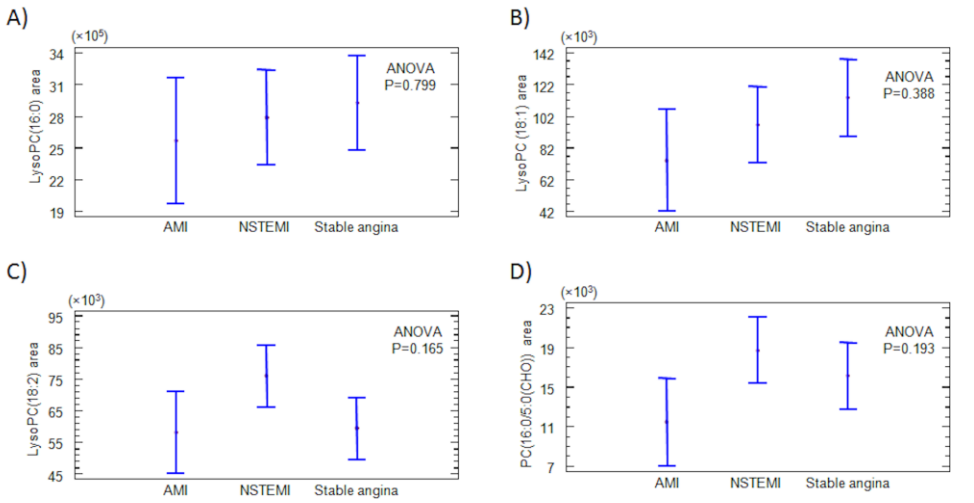
Supplementary figure 3. PCA scores plot obtained for the three groups of patients (AMI, NSTEMI and stable angina)



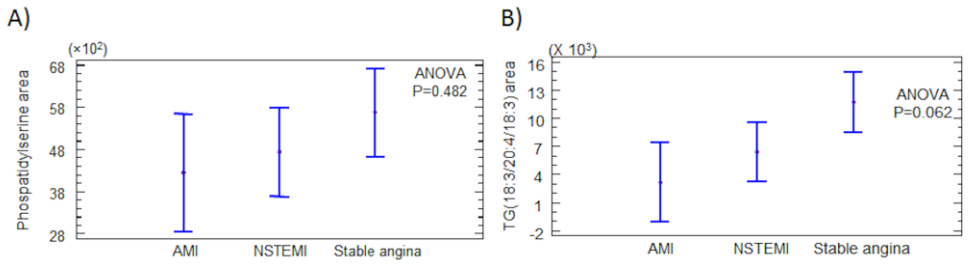
Supplementary figure 4. Means and 95% least significant difference (LSD) intervals for patients with hypercholesterolemia and those with normal cholesterol level for metabolites cholic acid (**A**) and deoxycholic acid (**B**).



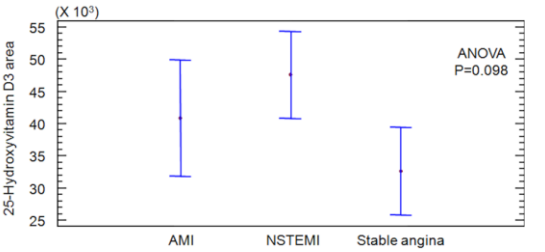
Supplementary figure 5. Means and 95% least significance difference (LSD) intervals for each group of patients for the metabolites cholic acid (**A**) and deoxycholic acid (**B**). *p*-Value of ANOVA test for all metabolites, and the cardiovascular disease as categorical factor are also shown.



Supplementary figure 6. Means and 95% least significance difference (LSD) intervals for each group of patients for the metabolites LysoPC(16:0) (**A**), LysoPC (18:1) (**B**), LysoPC (18:2) (**C**) and PC (16:0/5:0(CHO)) (**D**). *p*-Vvalue of ANOVA test for all metabolites, and the cardiovascular disease as categorical factor are also shown.



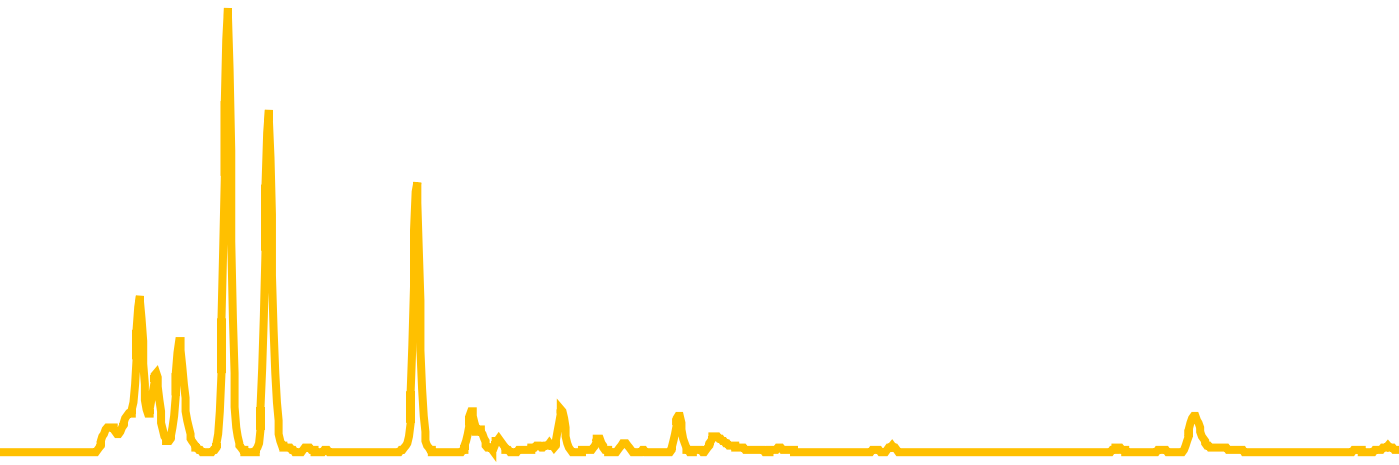
Supplementary figure 7.TG(18:3/20:4/18:3) peak area means and 95% least significance difference (LSD) intervals for each group of patients. *p*-Value of ANOVA test for this metabolite, using the cardiovascular diagnostic as categorical factor is also shown.

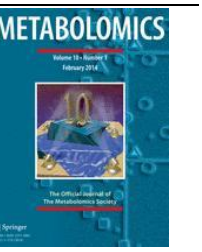


Supplementary figure 8. Hydroxyvitamin D3 peak area means and 95% least significance difference (LSD) intervals for each group of patients. *p*-Value of ANOVA test for this metabolite using the cardiovascular diagnostic as categorical factor is also shown.

Chapter 6:

Metabolomic markers to monitor atherosclerotic patients for myocardial infarction prevention





Metabolomic markers to monitor atherosclerotic patients for myocardial infarction prevention

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b]*, *N. Turck* [c], *X. Robin* [c], *J. G. Galache-Osuna* [d], *J. C. Sanchez* [c], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Translational Biomarker Group, Centre Médical Universitaire, Faculty of Medicine, Geneva University, Geneva, Switzerland*

[d]*Department of Cardiology, Hemodynamic and Interventional Cardiology, University Hospital Miguel Servet, Zaragoza, Spain*

Metabolomic markers to monitor atherosclerotic patients for myocardial infarction prevention

M. Calderón-Santiago, F. Priego-Capote, N. Turck, X. Robin, J. G. Galache-Osuna, J. C. Sanchez, M. D. Luque de Castro**

Abstract

Atherosclerosis is one of the most frequent aetiology for myocardial infarction and death. Therefore, the early prediction of an ischemic event can lead to a drastic reduction of mortality. In this research, serum marker metabolites from atherosclerotic patients who either had suffered from acute myocardial infarction or had been diagnosed with stable angina were analyzed by LC–QTOF MS/MS. Clinical factors such as age of the patients and body mass index (BMI), and biochemical factors such as urine levels of urea and creatinine were also considered as potential markers for the development of discriminative models. Individual markers with discrimination capability between individuals affected by myocardial infarct and those diagnosed by stable angina have been used to create a three-variable panel for infarct prediction. The panel consists of two lipid metabolites: 13-HpODE (a metabolite connected to the inflammatory cascade), lysophosphatidylcholine PC(22:6) (with an essential omega-3 fatty acid, docosahexanoic acid) and BMI. Sensitivity and specificity values were 85.1 and 80.8%, respectively, while positive and negative predictive values were 75 and 88.9%. Monitoring these markers could allow the early detection and risk stratification of myocardial infarct episodes with subsequent reduction of cardiovascular damages.

1. Introduction

One of the most frequent causes of death worldwide is related to cardiovascular diseases (CVDs) with over 80% of deaths in low- and middle-income countries and almost 50% of deaths in Europe (Petersen et al. 2005). CVDs are also a major cause of disability that significantly contributes to increase health care costs (Gaziano et al. 2006). These diseases usually have their origin in atherosclerosis, a coronary artery disease (CAD) that occurs when lipid plaques formed along the artery walls cause a subsequent thickening effect, which results in partial artery stenosis. In this stage, the symptomatic diagnostic is called angina. Then, the formation of thrombi (blood clots) on plaques surface can cause total occlusion of the artery leading to insufficient supply of oxygen and substrates to the cells (ischemia) that is produced as acute myocardial infarction (AMI).

Early diagnosis and risk stratification of myocardial infarction can lead to a reduction of mortality by initiation of the reperfusion treatment. Nowadays, this cardiovascular episode can be clinically diagnosed by an electrocardiogram (ECG), the main limitation of which is its alteration by other cardiopathies such as acute pericarditis or left ventricular hypertrophy resulting in false positives (Wang et al. 2003). Apart from ECG, measurement of some markers like creatine kinase isoenzymes, cardiac myoglobin, Human Fatty Acid Binding Protein (H-FABP) and troponins can help in the detection of AMI and risk stratification (Wu et al. 1996). Among them, troponins have been widely considered as the best cardiac markers, but conventional troponin assays lack sensitivity and precision at the low serum concentrations observed in the early hours after the onset of chest pain (Baker et al. 2011).The new ultra sensitive troponin assays may soon offer improved certainty in this early period with capability of reliable detection in patients with chronic stable heart disease and even in asymptomatic healthy individuals (Omland et al. 2009). One of the main limitations in using troponins as AMI markers is that acute and chronic conditions distinct from acute coronary syndromes (ACS) commonly lead to small elevations in troponin levels, with few available data regarding management of care for patients under such conditions

(de Lemos, 2013). Therefore, new statistical tools combined with highly sensitive assays, incorporating baseline troponin values and changes in concentration over 1 to 2 h may allow rapid exclusion of myocardial infarction and help to address specificity concerns, but they must be validated in appropriate target cohorts (Cullen et al. 2013).

Currently research on cardiac markers is also focused on the transcriptome and metabolome levels. Thus, the study of cardiac-specific microRNAs, one of the most active research areas, has resulted in a new marker, miR-208a, with a prediction model characterized by 90.9% sensitivity and 100% specificity (Devaux et al. 2012). However, additional investigations with large cohorts need to be done to validate microRNAs as cardiac markers and obtain the false positive rate. Furthermore, a more exhaustive research is required to make rapid detection of microRNAs possible and support their viability (Wang et al. 2010).

Metabolomics markers have also been proposed to study atherosclerosis events. Wang *et al.* identified choline, betaine,trimethylamine N-oxide, aconitic acid, threonine or hypoxanthine as metabolites with statistical relevance to aid in the atherosclerotic diagnostics (Wang et al. 2011).In other recent study, Shah *et al.* identified metabolite profiles (medium-chain acylcarnitines, short-chain dicarboxylacylcarnitines, long-chain dicarboxylacylcarnitines, branched-chain amino acids and fatty acids) for independent prediction of death/myocardial infarction (Shah et al. 2012). This study provided an opportunity to refine cardiovascular risk assessment and elucidate novel mechanisms of the disease. The specificity and sensitivity of these markers have been determined in only a short number of cases, but they have not been evaluated in multimetabolite panels.

The aims of this research were to find and configure panels of markers to increase the predictive power of myocardial infarction. Statistical models supported the selection of markers with discrimination capability between a group of atherosclerotic patients diagnosed with stable angina and another group affected

by acute myocardial infarct (AMI). The main risk factors considered in atherosclerosis and AMI were combined with metabolite markers as a strategy to evaluate their predictive capability either independently or in multiparametric diagnostic panels.

2. Materials and methods

2.1. Reagents

LC-grade methanol and MS-grade formic acid and acetonitrile (ACN) were from Scharlab (Barcelona, Spain). Deionized water (18 mΩ • cm) was from a Millipore Milli-Q water purification system.

2.2. Instruments and apparatus

A Sorvall Legend Micro 21R centrifuge (Thermo Scientific, Waltham, MA, USA) was used to centrifuge samples after deproteinization.

An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF tandem mass spectrometer equipped with a Jetstream® electrospray (ESI) source (Santa Clara, CA, USA) was used. The chromatographic eluate was monitored in high-resolution mode (2 GHz).

2.3. Cohort selected for the study

A total of 73 patients diagnosed by significant CAD were included in the study. CAD was diagnosed after evaluation through cardiac catheterization with a reduction of the arterial lumen ≥ 70% by angiographic stenosis. The main characteristics of the patients are shown in Table 1. As can be seen, the selected cohort was composed of 47 individuals diagnosed with stable angina and 26 individuals who had previously been affected by AMI (around 3 days prior to catheterization). The average age and body mass index (BMI) of the cohort was 66±13 years old and 28±4 kg/m^2, respectively. The cohort was composed by 86.3% male individuals and 50.7% smokers. Concerning clinical parameters, the CAD cohort was formed by 23.3% diabetic, 38.4% obese, 62.0% with

hypertension and 43.8% with hypercholesterolemia. These distributions were equilibrated in the two groups considered in this study: stable angina and AMI patients. Independence of these clinical factorsas well as smoking habit was proved according to the Pearson's Chi Squared test for categorical variables with a 95% of confidence level (*p*-values are shown in Table 1). Non-categorical factors were included for statistical analysis with potential metabolite markers.

Table 1 Main features of the cohort selected for this study. Independence of categorical variables was checked by P-Pearson Chi Squared test.

Characteristic	**Atherosclerosis patients (n=73)**	**Patients with AMI (n=26)**	**Patients with stable angina (n=47)**	***p*-value**
Age (mean± SD, years)	66±13	68±12	65±13	-
Male, n (%)	63 (86.3)	23 (88.5)	40 (85.1)	0.690
Smoking, n (%)	37 (50.7)	12 (46.1)	25 (53.2)	0.565
Diabetes, n (%)	17 (23.3)	8 (30.7)	9 (19.1)	0.278
Obesity, n (%)	28 (38.4)	9 (34.6)	19 (40.4)	0.436
Hypertension, n (%)	44 (62.0)	14 (53.8)	30 (66.7)	0.284
Hypercholesterolemia, n (%)	32 (43.8)	11 (42.3)	21 (44.7)	0.845
BMI (kg/m^2)	28±4	27±4	29±3	-

Clinical examination was completed by quantitative analysis of urea and creatinine levels in urine following conventional protocols (using the auto-analyzers AU2700 and AU5400 from Beckman Coulter, Marseille, France). Most patients included in the cohort were under controlled medication as follows: β-blocking agents (78%), statins (92%), antiplatelet agents (88%) and aspirin (98%). Therefore, drugs administration was not considered in this study.

2.4. Blood drawing and serum isolation

Blood drawing immediately preceded catheterization. All steps from blood to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004), which were supervised by specialized personnel from Miguel Servet Hospital (Zaragoza, Spain) that approved the experiments. Individuals selected for this study were informed to obtain consent prior to sample collection.

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated for 30 min at room temperature to allow coagulation. Then, the tubes were centrifuged at 2000 *g* for 15 min at 4 °C to isolate the serum fraction that was placed in plastic tubes and stored at −80 °C until analysis.

2.5. Sample treatment

The vials containing the serum samples (100 μL) were immersed in an ice bath and treated for deproteinization with methanol (200 μL) (Bruce et al. 2009). The mixture was shaken for 1 min and the precipitate removed after centrifugation for 5 min at 4 °C and 13800 *g*. The upper phase was collected in a vial, then placed in the autosampler of the chromatograph for subsequent analysis.

2.6. LC–QTOF MS/MS analysis

Chromatographic separation was performed using a Mediterranea reversed phase C18 analytical column (100 mm×0.46 mm i.d., 3 μm particle size) from Teknokroma (Barcelona, Spain) kept at 25 °C. Mobile phases were water (phase A) and ACN (phase B) both with 0.1% formic acid as ionization agent. The LC pump was programmed at a flow rate of 0.8 mL/min and the following gradient elution was carried out: 3% phase B was kept constant from min 0 to 2; from 3 to 100% of phase B since min 2 to 27; finally, 100% of phase B was kept since min 27 to 33.

The injection volume was 10 μL and the injector needle was rinsed with 70% methanol for 10 times between injections. Furthermore, the needle seat back was flushed for 15 s at a flow rate of 4 mL/min with 70% methanol to avoid cross contamination. Electrospray source parameters, operating in both positive and negative ionization mode, were as follows: nozzle, capillary and fragmentor voltage were set at ±2 kV, ±3.5 kV and 175 V, respectively; N_2 nebulizer gas was flowed at 40 psi; N_2 drying gas flow rate and temperature were 8 L/min and 325 °C, respectively; N_2 sheath gas flow rate and temperature were 11 L/min and 350 °C, respectively. The instrument was calibrated and tuned according to

procedures recommended by the manufacturer. The data were collected in both centroid and profile modes at a rate of 2 spectra per second. Firstly, the samples were analyzed in full scan MS mode in the *m/z* range 60–1100 for metabolite profiling. The reported data set was statistically processed to detect molecular features with a high significance to explain the observed variability. Once the most significant entities were recognized, the samples were analyzed in targeted MS/MS mode (*m/z* range 60–1100) to obtain structural information with identification purposes at three collision energies 10, 20 and 40 eV. To assure the desired mass accuracy of recorded ions, continuous internal calibration was performed during analyses by using the signals at *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct of hexakis (1H, 1H, 3H-tetrafluoropropoxy)phosphazine or HP-921) in negative ionization mode; while in positive ion mode ions with *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated HP-921] were used. Resolution provided by the instrument ranged from 15000 FWHM (Full Width at Half Maximum) at low masses (purine) and 30000 FWHM at high mass values (HP-921).

2.7. Data processing and statistical analysis

Raw data files were converted to mzData files using MassHunter Workstation software (version 3.01 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) and then processed in R statistical language (version 2.15.0, http://www.r-project.org/) using the two open-free XCMS R-package (version 1.24, http://metlin.scripps.edu/xcms/index.php) (Carpenter and Bithell, 2000; Robin et al. 2011) and CAMERA R-package (version 1.12.0, http://www.bioconductor.org/packages/2.10/bioc/html/CAMERA.html). First, the XCMS package was used for processing mzData files by extracting potential molecular features (MFs) considering only ions exceeding 1000 counts peak height with a peak width between 10 and 60 s a signal-to-noise threshold of 10 and 5 ppm of error in mass accuracy. Secondly, a non-linear alignment was executed using the "retcor" function with a degree of smoothing for local polynomial regression fitting of 0.5 and a symmetric fitting, process which used a re-descending M estimator with Tukey's biweight function. Then, the CAMERA

package was used to correlate potential adducts ($[M-H]^-$, $[M+Cl]^-$ and $[M+HCOOH-H]^-$ in negative ionization mode; and $[M+H]^+$ and $[M+Na]^+$ in positive ionization mode) and isotopic peaks of the same molecular entity. Ions with identical elution profiles and related *m/z* values (representing different adducts or isotopes of the same compound) were grouped as a unique feature to remove redundant information. This process resulted in a data set for each polarization mode containing the intensity values in the apex of chromatographic peaks for all molecular entities characterized by accurate mass and retention time (RT). Background contribution was removed by subtraction of MFs linked to plasticizers, solvent impurities and other contaminating compounds after analysis of blank sample (methanol) under identical instrument operation conditions. Data sets were then imported to the Mass Profiler Professional (MPP) software package (version 2.0, Agilent Technologies, Santa Clara, CA, USA) for further processing. Data pretreatment was based on baselining to median across samples to remove background noise and normalization by logarithmic transformation to reduce relatively large differences among MF abundances. A fold-change algorithm was applied as filter to retain those molecular entities, which experienced a change in relative concentration of 1.75 between the two classes. Significance of the remaining entities was studied by Kruskal-Wallis test, after verifying their non-normal distribution by Kolmogorov-Smirnov test.

MPP software also allowed unsupervised analysis of the data set by PCA. Mean centering as pretreatment scaled the data set.

The p-ROC R-package (1.5.1 version, URL http://web.expasy.org/pROC/) was used to build ROC curves as statistical tool to study the influence of risk factors related to CAD. This study was carried out by using multiparametric pROC to differentiate groups formed by combination of diagnostic (stable angina or AMI) and risk factor with categorical *yes/no* response. Thresholds of individual predictors were set to provide specificity above 90% considering that an efficient predictor in clinical practice should be able to detect clearly at least nine out of ten patients having a favorable prognosis when the test was negative. Partial ROC AUCs (pAUC) (Fawcett, 2006; Robin et al. 2011) were calculated

using an adaptation of algorithms previously reported (Carpenter and Bithell 2000). PanelomiX toolbox, which is supported on the iterative combination of biomarkers and thresholds, was used to combine biomarkers and other scores by selecting thresholds that provide optimal classification performance in terms of sensitivity and specificity (Robin et al. 2013). Panels of markers were built by setting sensitivity and selectivity cut-off values >80%.

Identification of the most relevant entities was done using MS and MS/MS information and searching in the METLIN MS and MS/MS database (http://metlin.scripps.edu) and Human Metabolome Database (HMDB, v. 3.5).

3. Results and discussion

3.1. Data pre-treatment and statistical analysis

Serum samples were analyzed by LC–QTOF MS/MS both in positive and negative ionization modes. Nevertheless, the data set obtained in the positive ionization mode did not report significant differences between the two groups under study. For this reason, this section is entirely focused on the results obtained by LC–QTOF MS/MS in negative ionization mode.

As previously mentioned, MFs were extracted from the raw data files obtained after LC–TOF MS analysis of serum taking into account alignment of chromatograms to correct small differences in retention time and mass accuracy. The resulting data set was composed by 73 samples×289 molecular entities, 92.7% of which were detected in at least 90% of the total number of samples. This fact ensures representativeness of the study since no molecular entities associated to particular individuals, which could divert the resulting statistical models, were included in the data set. Once the data set was built, a fold change algorithm with cut-off 1.75 was applied to reduce the data set to those molecular entities that exhibited a change of relative concentration ascribed to both CAD events. The application of this filter reduced the number of entities up to 26 tentative metabolites defined by the pair *m/z* value and retention time.

This data set was completed by clinical parameters obtained after routine urine analysis which reported information on levels of creatinine and urea. Both parameters have shown to be linked to risk cardiovascular factors (Alderman et al. 1995; Kapitulnik 2004). Anthropometric information through body mass index (BMI), expressed as kg/m^2, and patients age were also included in the data set, which was finally formed by 30 variables (26 potential metabolites, BMI, age, creatinine and urea).The Kolmogorov-Smirnov test was used to check if these 30 variables showed normal distributions in the cohort under study. As these variables did not show a normal distribution, the non-parametric Kruskal-Wallis test was applied to identify the most influential variables to explain the CAD diagnostic. Thirteen molecular features and the BMI reported significance with *p*-value below 0.1, which was selected to keep a suited number of variables for multivariate analysis. Supplemental Table S1 shows the values of parameters describing the significance of variables. These 14 variables (13 MFs and BMI) formed a new data set that was analyzed by principal component analysis (PCA) to find discrimination trends according to the CAD diagnosis: stable angina or AMI. As shown by the PCA plots (Fig. 1), there is a discrimination trend that gives partial separation between patients diagnosed with stable angina and those who had suffered an AMI episode. Separation was visualized by plotting PC1, PC2 and PC3 components that explained 61.6% of the total variability.

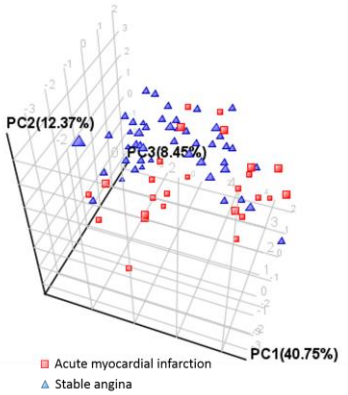


Figure 1. PCA scores plot to find discrimination patterns between CAD patients diagnosed with stable angina or after an AMI episode.

3.2. Influence of risk factors in the diagnosis of CAD events

There are numerous risk factors associated to cardiovascular diseases. Among them, it is worth mentioning diabetes, obesity, hypertension, smoking habit, hypercholesterolemia, lack of physical activity, unhealthy diet, stress or genetic factors. In this study, obesity, hypercholesterolemia and smoking habit were selected to evaluate their prediction capability. For this purpose, statistical analysis by the Kruskal-Wallis test raising significance with p-value<0.05 as cut-off was applied to identify the most influential metabolites contributing to differentiate both groups of individuals (Table 2). Five molecular features were detected as highly significant together with the BMI. Search of MS/MS spectra in databases led to the identification of biliverdin and bilirubin, two eicosanoids (5-HETE and 13-HpODE) and the 1Z-alkenylacylglycerolDG (P-14:0/18:1n9). Bilirubin is an excretion product formed in normal heme catabolism, which can be found in human serum (Kapitulnik 2004). However, biliverdin is an intermediate formed in the same pathway but rapidly reduced to bilirubin by the activity of biliverdin reductase. This metabolite has not been detected in human serum and, therefore, its presence could be justified by partial oxidation of bilirubin from sampling to analysis. Concerning 5-HETE and 13-HpODE metabolites, their identification was supported on the selective product ions formed by MS/MS fragmentation, which allows discriminating them from their structural isomers (Ferreiro-Vera et al. 2011). The fifth metabolite, 1Z-alkenyl-acylglycerol DG (P-14:0/18:1n9), is an intermediate in the biosynthesis of ether phospholipids, which are known as plasmalogens. These metabolites have been detected in cardiovascular tissues where they are supposed to exert antioxidative functions as scavengers of reactive oxygen species (Stenvinkel et al. 2004). Plasmalogen oxidation products as α-hydroxyaldehydes and epoxides accumulate in all chronic diseases as atherosclerosis and myocardial infarction as well as in different neuropathologies and aging. Table 2 shows the information used for identification of the target metabolites with the highest discrimination capability.

Table 2. List of identified metabolites including retention time (RT), monoisotopic molecular weight (M_w), *m/z* value of the detected adducts and product ions and the error expressed in ppm.

Metabolite	M_w	RT (min)	Parent ions (*m/z*)	Product ions (*m/z*)	Δppm
Bilirubin	584.2818	31.5	605.2380 $[M+Na-2H]^-$ 583.2565 $[M-H]^-$	285.1301	0.2
Biliverdin	582.2478	31.6	581.2405 $[M-H]^-$	358.7100	0.1
13-HpODE	312.2300	29.4	357.2263 $[M+FA-H]^-$	311.1689 293.1795 183.0120	5.0
5-HETE	320.2351	23.3	319.2289 $[M-H]^-$	115.0396	3.0
DG(P-14:0/18:1n9)	550.4961	29.4	595.4904 $[M+FA-H]^-$	225.0080 281.2461 535.0597	6.0

Once identification of potential metabolites was completed, the next step was to study their prediction capability to discriminate the CAD diagnostic, stable angina or AMI in conjunction to the three risk factors: obesity, smoking and hypercholesterolemia. Biliverdin was excluded from this study since this metabolite seems to be produced by degradation of bilirubin. The study was carried out by using multiparametric pROC, as previously described (specificity above 90%). Under these conditions, the maximum partial area under the curve (pAUC) for a ROC curve was 10%. Table 3 lists the main parameters and performance of each metabolite fulfilling the specificity requirement, while Fig. 2 shows the prediction performance (through ROC curves) of each metabolite for discrimination of the different indication/risk factor combinations. Smoking and obesity were the two risk factors that led to prediction capabilities with specificity above 90% and sensitivity values above 40% for DG(P-14:0/18:1n9), 13-HpODE and BMI. Particularly, the inclusion of obesity led to sensitivity values around 44% for DG(P-14:0/18:1n9) and 13-HpODE, while the BMI pROC reported 58% for the same parameter. On the other hand, sensitivity reached 50% for the same two metabolites when smoking habit was considered. This fact suggested the evaluation of the prediction capability of the target metabolites by selecting only

patients defined by the risk factors: obese, hypercholesterolemic and smokers. In this context, the best prediction capability was for hypercholesterolemic patients, for which sensitivity values reached 48% for BMI, 57% for bilirubin and 5-HETE and 62% for 13-HpODE. Therefore, it is clear that the prediction capability of individual markers improved when hypercholesterolemic patients were evaluated. In the case of obese patients and smokers, the prediction capability did not lead to better sensitivity as compared to models including patients non defined by risk factors.

Table 3. Parameters of the ROC curves obtained for each identified metabolite by selecting a specificity range from 90 to 100% considering classes generated after combining cardiovascular disease (CVD) with risk factors (smoking habit (SH), obesity and hypercholesterolemia (HC)). Sensitivity, specificity and partial area under the curve (pAUC) parameters are expressed as percent. Curves reaching a sensitivity value higher than 40% are highlighted in grey.

Compound	Parameter	CVD	CVD + SH (4 classes)	CVD in smokers	CVD + Obesity (4 classes)	CVD in obese patients	CVD + HC (4 classes)	CVD in patients with HC
Bilirubin	Sensitivity	38	20	0	28	0	24	57
	Specificity	92	96	100	93	100	92	91
	pAUC	0.8	1.2	0.0	2.1	0.0	2.2	4.8
13-HpODE	Sensitivity	28	50	32	42	48	25	62
	Specificity	93	91	92	97	100	100	91
	pAUC	2.2	1.1	2.9	3.1	4.7	2.7	4.9
DG(P-14:0/18:1n9)	Sensitivity	28	50	42	42	31	26	32
	Specificity	93	91	92	96	100	100	91
	pAUC	1.7	0.5	1.4	2.6	3.2	2.7	2.9
BMI	Sensitivity	18	21	22	58	6	6	48
	Specificity	92	91	92	93	100	98	91
	pAUC	0.6	1.9	0.7	3.9	3.7	0.3	4.3
5-HETE	Sensitivity	20	8	3	11	38	0	57
	Specificity	92	91	92	93	100	100	91
	pAUC	0.8	0.1	0.1	0.7	0.5	0.0	4.4

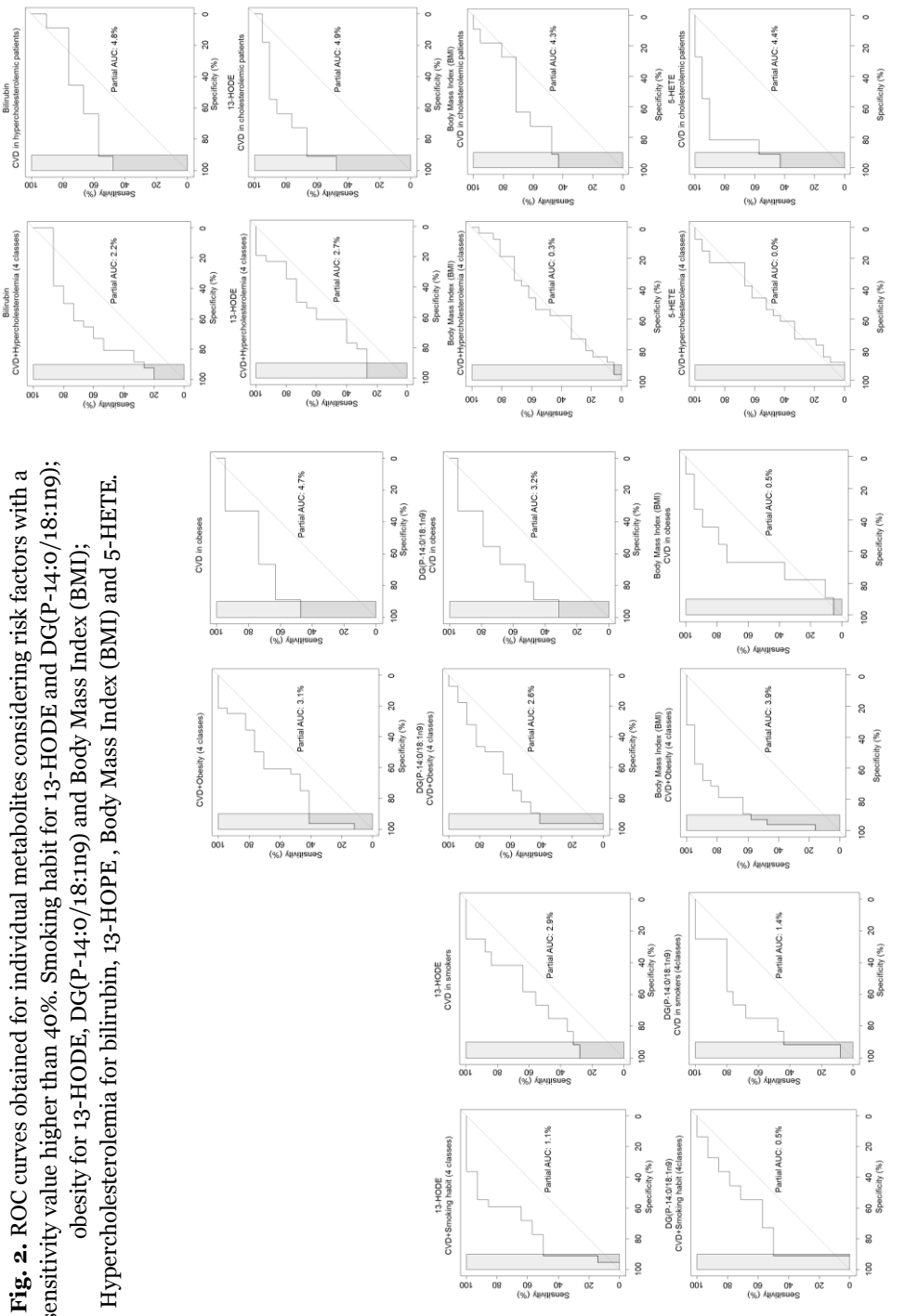


Fig. 2. ROC curves obtained for individual metabolites considering risk factors with a sensitivity value higher than 40%. Smoking habit for 13-HODE and DG(P-14:0/18:1n9); obesity for 13-HODE, DG(P-14:0/18:1n9) and Body Mass Index (BMI); Hypercholesterolemia for bilirubin, 13-HOPE , Body Mass Index (BMI) and 5-HETE.

3.3. Panel development

The development of a panel for discrimination between CAD patients diagnosed with stable angina and those affected by AMI was addressed using PanelomiX as computational toolbox. In this research, the initial data set was formed by 30 variables (26 metabolites and BMI, urea and creatinine levels and age). The configuration of the panel was supported on two statistical criteria: at least 80% sensitivity and at least 80% specificity. Thus, both false negative and positive rates would be considerably decreased. PanelomiX was applied to the cohort formed by 73 CAD patients to define a panel of three markers with the best classification parameters. Thereby, the best panel providing specificity and sensitivity above 80% was obtained with two variables corresponding to two metabolites and the BMI. This panel enabled a model with 85.1% specificity and 80.8% sensitivity, while the pAUC at 95% CI was 9.7 (5.7–15.7)%. Furthermore, positive and negative predictive values were 75 and 88.9%, respectively. One of the metabolites involved in this panel was the 13-HpODE, which was previously detected as a highly significant metabolite to discriminate between atherosclerotic patients. The second metabolite was identified as lysophosphatidylcholine, particularly lysoPC(22:6). Identification of this metabolite is described in Fig. 3 that contains MS/MS spectra in positive and negative ionization modes. As can be seen, product ions from fragmentation of the phosphatidylcholine moiety were identified in positive ionization mode. On the other hand, the characteristic $[M–CH_3]^-$ fragment, which is representative of phosphatidylcholines, and the product ions corresponding to the acyl chain $R\text{-}COO^-$ (C22:6) and its decarboxylated fragment were identified in negative mode. This panel formed by two lipids and BMI is especially interesting since it combines two lipidic metabolites with an anthropometric variable directly related to obesity. Furthermore, the lipids included in this panel belong to two families that have been related to cardiovascular diseases through oxidative stress. Firstly, 13-HpODE, one of the primary oxidation products of linoleic acid, is a significant component of oxidized LDL, which is implicated in the pathogenesis of atherosclerosis. The concentration of 13-HpODE has previously been linked to

atherosclerotic lesions (Natarajan et al. 2001). Secondly, anomalies in levels of lysophosphatidylcholines have been shown to be involved in cardiovascular complications (Matsumoto et al. 2007).

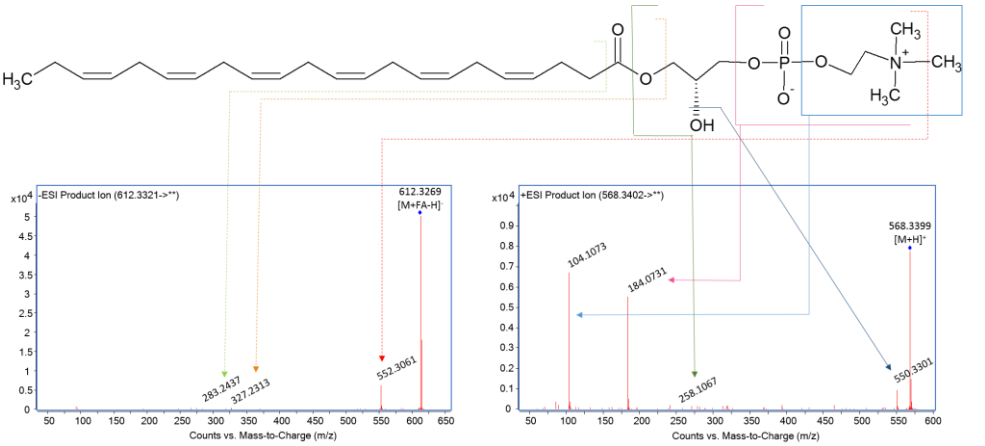


Fig. 3 Fragmentation pattern and MS/MS spectra in positive and negative ionization modes for lysophosphatidylcholine PC(22:6).

4. Conclusions

The analysis of serum samples by LC–QqTOF has shown to be a suitable strategy for biomarker discovery. The influence of risk factors has been studied by analyzing the prediction capability of potential markers in patients with a specific risk factor. In this context, a better prediction capability was obtained for hypercholesterolemic patients, for which specificity values of 91% and sensitivity values from 48 to 62% were obtained for bilirubin, 13-HpODE, BMI and 5-HETE.

A panel of markers for myocardial infarction prediction have been proposed with a specificity of 85.1 % and a sensitivity of 80.8 %, which gave rise to low false negative and positive rates. The metabolites that composed this panel are two lipids, 13-HpODE and lysoPC(22:6), and BMI, an anthropometric variable directly related to obesity. These results warrant a larger scale study to validate the proposed model.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2012-37428). F. Priego-Capote and M. Calderón Santiago are also grateful to the Spanish Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921) and an FPU scholarship (AP2009-0499), respectively. The Miguel Servet Hospital (Zaragoza, Spain) is also thanked for donors' serum samples.

References

Alderman, M.H., Madhavan, S., Cohen, H., Sealey, J.E., & Laragh, J.H. (1995). Low urinary sodium is associated with greater risk of Myocardial Infarction among treated hypertensive men. *Hypertension*, 25, 1144–1152.

Baker, J.O., Reinhold, J., Redwood, S., & Marber, M.S. (2011). Troponins: redefining their limits. *Heart,* 97(6), 447–452.

Bruce, S.J., Tavazzi, I., Parisod, V., Rezzi, S., Kochhar, S., & Guy, P.A. (2009). Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Analytical Chemistry,* 81 (9), 3285–3296.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: when, which,what? A practical guide for medical statisticians. *Statistics in Medicine,* 19, 1141–1164.

Cullen, L., Mueller, C., Parsonage, W.A., et al. (2013). Validation of high-sensitivity Troponin I in a 2-hour diagnostic strategy to assess 30-day outcomes in Emergency Department patients with possible acute coronary syndrome. *Journal of the American College of Cardiology,* 62(14), 1242–1249.

De Lemos, J.A. (2013). Increasingly sensitive assays for cardiac troponins: a review. *JAMA – Journal of American Medical Association,* 309(21), 2262–2269.

Devaux, Y., Vausort, M., Goretti, E., et al. (2012). Use of circulating microRNAs to diagnose acute myocardial infarction. *Clinical Chemistry,* 58(3), 559–567.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters,* 27, 861–874.

Ferreiro-Vera, C., Mata-Granados, J.M., Priego-Capote, F, Quesada-Gómez, J.M., & Luque de Castro, M.D. (2011). Automated targeting analysis of eicosanoid inflammation biomarkers in human serum and in the exometabolome of stem cells by SPE–LC–MS/MS. *Analytical and Bioanalytical Chemistry,* 399, 1093–1103.

Gaziano, T., Reddy, K.S. , Paccaud, F. , Horton, S., & Chaturvedi, V. (2006) Cardiovascular Disease in Jamison, D.T., Breman, J.G., Measham, A.R., et al. (Eds), Disease Control Priorities in Developing Countries. World Bank: Washington (DG), U.S.A., pp. 645–62.

Kapitulnik, J. (2004). Bilirubin: an endogenous product of Heme degradation with both cytotoxic and cytoprotective properties. *Molecular Pharmacology,* 66(4), 773–779.

Matsumoto, T., Kobayashi, T., & Kamata, K. (2007). Role of lysophosphatidylcholine (LPC) in atherosclerosis. *Current Medicinal Chemistry,*14, 3209–3220.

Natarajan, R., Reddy, M.A., Malik, K.U., Fatima, S., & Khan, B.V. (2001). Signaling mechanisms of nuclear factor-κB-mediated activation of inflammatory genes by 13-hydroperoxyoctadecadienoic acid in cultured vascular smooth muscle cells. *Arteriosclerosis, Thrombosis, and Vascular Biology,* 21, 1408–1413.

Omland, T., de Lemos, J.A., Sabatine, M.S., et al. (2009). A sensitive cardiac troponin T assay in stable coronary artery disease. *The New England Journal of Medicine,* 361, 2538–2547.

Petersen, S., Peto, V., Rayner, M., Leal, J., Luengo-Fernández, R., & Gray, A. (2005) European Cardiovascular Disease Statistics, British Heart Foundation, London, U.K.

Robin, X., Turck, N., Hainard, A., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics,* 12, 77–84.

Robin, X., Turck, N., Hainard, A., et al. (2013). PanelomiX: a threshold-based algorithm to create panels of biomarkers. *Translational Proteomics,* 1(1), 57–64.

Shah, S.H., Sun, J.L., Stevens, R.D., et al. (2012). Baseline metabolomic profiles predict cardiovascular events in patients at risk for coronary artery disease. *American Heart Journal,* 163(5), 844–850.

Stenvinkel, P, Diczfalusy, U., Lindholm, B., & Heimbürger, O. (2004). Phospholipid plasmalogen, a surrogate marker of oxidative stress, is associated with increased cardiovascular mortality in patients on renal replacement therapy. *Nephrology Dialysis Transplantation,* 19, 972–976.

Wang, G.K., Zhu, J.Q., Zhang, J.T., et al. (2010). Circulating microRNA: a novel potential biomarker for early diagnosis of acute myocardial infarction in humans. *European Heart Journal,* 31(6), 659–666.

Wang, K., Asinger, R.W., & Marriott, H.J. (2003). ST-segment elevation in conditions other than acute myocardial infarction. *The New England Journal of Medicine,* 349, 2128–2135.

Wang, Z., Klipfell, E., Bennett, B.J., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature,* 472, 57–63.

Wannamethee, G., Shaper, A.G., & Perry, I.J. (1997). Serum creatinine concentration and risk of cardiovascular disease a possible marker for increased risk of strokes. *Stroke,* 28, 557–563.

Wu, A.H., Feng, Y.J., Contois, J.H., & Pervaiz, S. (1996). Comparison of myoglobin, creatine kinase-MB, and cardiac troponin I for diagnosis of acute myocardial infarction. *Annals of Clinical and Laboratory Science,* 26(4), 291–300.

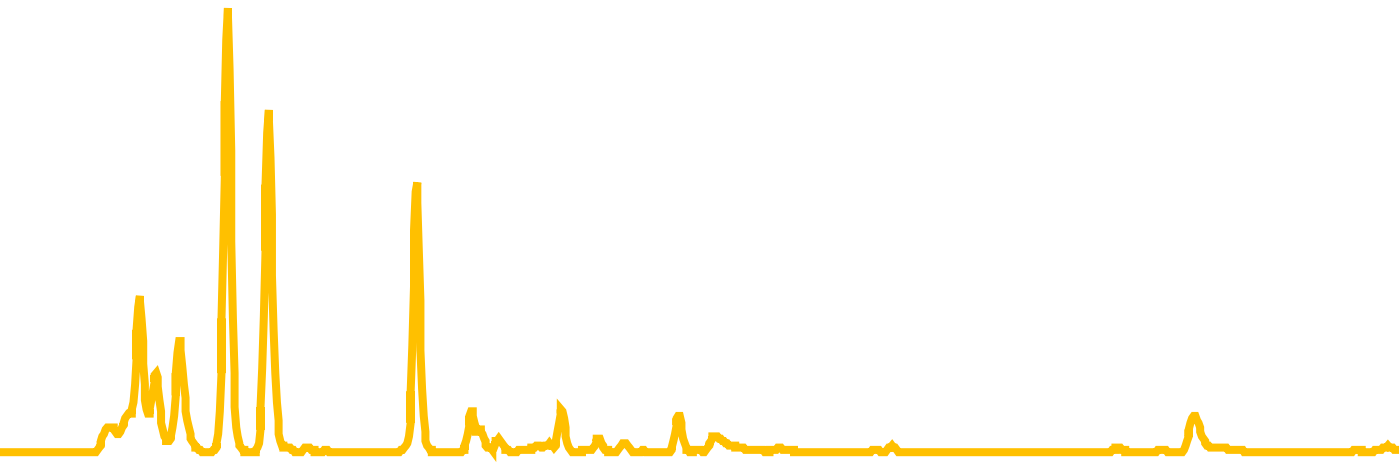
Supplementary material

Supplementary Table 1. Statistical p-value obtained after application of the Kruskall-Wallis test considering the cardiovascular atherosclerotic disease diagnosis as dependent variable and individual potential metabolites or clinical factors as independent variables. Only those variables with p-value below 0.10 are shown.

Independent variable	p-value
357.2263	0.003
595.4904	0.004
Body Mass Index (BMI)	0.006
581.2405	0.007
605.2380	0.008
319.2289	0.036
804.5745	0.051
381.1743	0.056
105.9603	0.060
612.3289	0.087
383.1896	0.090
407.2794	0.090
293.1793	0.099
391.2851	0.099

Chapter 7:

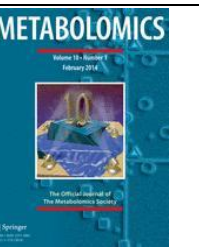
Human sweat metabolomics for lung cancer prediction



Springer

Sent to

Metabolomics



Human sweat metabolomics for lung cancer prediction

M. Calderón-Santiago [a,b], F. Priego-Capote [a,b], N. Turck [c], X. Robin [c], B. Jurado-Gámez [d], J. C. Sanchez [c], M. D. Luque de Castro[a,b]**

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Translational Biomarker Group, Centre Médical Universitaire, Faculty of Medicine, Geneva University, Geneva, Switzerland*

[d]*Department of Respiratory Medicine, Reina Sofía Hospital, University of Córdoba, Córdoba, Spain*

Human sweat metabolomics for lung cancer prediction

M. Calderón-Santiago, F. Priego-Capote, N. Turck, X. Robin, B. Jurado-Gámez, J. C. Sanchez, M. D. Luque de Castro**

Abstract

Lung cancer is the carcinogenic disease with the highest mortality rate owing to the advanced stage at which it is usually detected. For this reason, methods for lung cancer detection at early stages are needed. In this context a method based on the metabolite analysis of sweat to discriminate between patients with lung cancer versus control individuals (smoker and non-smoker individuals) is proposed. The capability of the metabolites identified in sweat to discriminate between both groups was studied. Among them, a trisaccharide phosphate presented the best independent performance in terms of the specificity/sensitivity pair (80 and 72.7%, respectively). Additionally, two panels of metabolites were also configured using the PanelomiX tool as an attempt to reduce false negatives (at least 80% specificity) and false positives (at least 80% sensitivity). The first panel (100% specificity and 63.6% sensitivity) was composed by nonanedioic acid, γ-GluLeu dipeptide and maltotriose, while the second panel (88.6% specificity and 81.8% sensitivity) included nonanedioic acid, maltotriose and the monoglyceride MG(22:2). As both panels were supported on high sensitivity and specificity values, the proposed approach would be based on the analysis of the four implicated metabolites. The combined use of the two panels would allow reducing the number of cases subjected to confirmatory tests with the minimum rate of false negative and positive rates (0% and 18.2%, respectively).

1. Introduction

Lung cancer is nowadays the neoplasic disease associated to the highest mortality worldwide. According to the American Cancer Society, lung cancer causes more deaths than the sum of the three most common cancers (colon, breast and prostate) (World Health Organization, 2012). Despite the advances in diagnostic, surgery and chemotherapy, overall survival rate 5 years after detection is below 15% (Howlander et al. 2010). The high mortality rate of lung cancer is mainly explained by the advanced stage at which it is usually detected as available diagnostic tests are invasive and, therefore, they cannot be used for general screening.

Presently, non-invasive techniques for cancer screening are the subject of wide clinical research. The use of diagnostic screening by sputum or thorax radiography has not succeeded in the reduction of mortality (Wegwarth et al. 2012). Recently, the utilization of techniques such as low-dose computed tomography and chest X-ray radiography have achieved to decrease mortality rate around 20% (National Lung Screening Trial Research Team, 2012). Nevertheless, the main limitation of the present methods to diagnose lung cancer is ascribed to the high number of abnormalities potentially mimicking lung cancer (false positives) and the difficulty in applying diagnostic tools to the target population because of the high costs.

The above limitations have driven active research in omics disciplines with two main purposes: to understand the development and biology of lung cancer and identify potential biomarkers to detect initial stages in the development of the disease (Khadir and Tiss, 2013). Specifically, recent proteomic research has led to identification of potential biomarkers of lung cancer in saliva (Xiao et al. 2012) or plasma (Fan et al. 2009). On the other hand, recent genomic and transcriptomic studies have been focused on the detection of mutations related to cancer, as an attempt to understand the mechanism of tumour generation and the influence of risk factors such as smoking habit (Girard et al. 2000; Pleasance et al. 2009; Beane et al. 2011).

Metabolomics has provided new technical approaches to enhance the capability of this discipline in biomarker discovery for clinical diagnosis, as many of the analytical tools used in metabolomics are automatable and, on the whole high throughput, they lend themselves to screening populations of individuals for disease. Recent research in this field has focused on biomarker discovery using different biological samples (tissue, urine, exhaled breath condensate or plasma) (Beger 2013). The most common strategy has been metabolic profiling, which has been applied to conventional biofluids such as urine and serum/plasma. However, the complexity of serum/plasma and urine matrices is a factor to be considered in the search for biomarkers since other pathological conditions could interfere, particularly when a panel of markers formed by metabolites involved in primary metabolism is found. This is the case with amino acids or compounds involved in energy metabolism (glycolysis or Krebs' cycle) (Kami et al. 2013) as they have shown to be modified in tumour cells. Concretely, potential markers found in urine included modified nucleosides, sarcosine, creatinine or acetylglutamine (Beger 2013) (Carrolla et al. 2011). However, the measurement of modified nucleosides levels have not shown to be reproducible or robust and sarcosine has also shown to fail in prostate cancer detection (Jentzmik et al. 2010). On the other hand, potential markers found in plasma/serum included mainly proteins and metabolites as lysophosphatidylcholines or phosphatidylinositols (Beger 2013), compounds previously related to other diseases (Karlsson et al. 2006).

Despite blood and urine are the most common human biofluids in clinical studies on lung cancer, other biofluids can be easily obtained and in a noninvasive manner. Sweat, saliva or tears, characterized by less complex matrices, could be evaluated in cohorts involving individuals affected by lung cancer (Kutyshenko et al. 2011). In contrast to the abundant research using serum/plasma or urine as target biofluids, sweat has received scant attention owing to the lack of homogenous sweat sampling protocols providing enough volume for analysis. However, the sensitivity of present available instrumentation makes possible obtaining metabolite profiles from small sample volumes (Wing

et al. 2013; Álvarez-Sánchez et al. 2010). Furthermore, sweat collection protocols have also been improved (Mena-Bravo and Luque de Castro 2014). In fact, there are already methods in which sweat is analyzed for disease discrimination such as in the analysis of chloride to diagnose cystic fibrosis (Gibson and Cooke 1959; Lynch 2010; Kirk et al. 1992). Sweat has also been used to detect the presence of toxics in the organism (Giovanni and Fucci 2013; Kintz et al. 1996; Taylor et al. 1998; Uemura et al. 2004; Sears et al. 2012) or to control the hydro–electrolytic relationship during exercise in humans (Lee et al. 2010; Hew-Butler et al. 2010). All these findings, together with present cutting-edge analytical technologies, have led the authors to study sweat as a potential sample to diagnose lung cancer.

The aim of the present study was to evaluate sweat to discriminate individuals affected by lung cancer by metabolomics analysis using liquid chromatography–tandem mass spectrometry in high-resolution mode (LC–QTOF MS/MS). Metabolite profiles obtained from diagnosed patients were compared to those provided by a control cohort and most significant metabolites identified were studied to evaluate their prediction capability. Statistical analysis of those metabolites supported the performance of a panel to reduce the rate of individuals to be subjected to invasive intervention for lung cancer diagnosis.

2. Materials and methods

2.1. Chemicals and reagents

MS-grade formic acid and acetonitrile (ACN) to prepare the chromatographic mobile phases were from Scharlab (Barcelona, Spain). Deionized water (18 mΩ • cm) from a Millipore Milli-Q water purification system was used.

2.2. Instruments and apparatus

An Agilent 1200 Series LC system (consisting of a binary pump, a vacuum degasser, an autosampler and a thermostated column compartment) coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped

with a dual electrospray ionization (ESI) source (Santa Clara, CA, USA) was used. The chromatographic eluate was monitored in high-resolution mode.

2.3. Cohort selected for the study

Forty one patients of the Respiratory Medicine Department diagnosed with lung cancer from May to December 2012 were included in this study. All patients with lung cancer were diagnosed after clinical tests based on bronchoscopy, fine-needle biopsy, or video-assisted thoracoscopy. These patients had an average age of 62±11 years and 78% of them were males. Their main characteristics —age, sex, body-mass index (BMI) and smoked pack-year— are shown in Supplementary Table 1. Independence of these factors among them and the control group was proved by analysis of variance (ANOVA) with a 95% of confidence level, except for the age that was proved by Kruskal-Wallis as it did not fit a normal distribution according to the Skewness and Kurtosis criteria for normality (*p*-values are shown in Supplementary Table 1).

The present accepted guidelines for pathological and staging diagnosis of lung cancer were used (Sánchez de Cos et al. 2011). The most frequent diagnostic was squamous cell carcinoma (21 patients, 51%), followed by adenocarcinoma (8 patients, 19.5%), small cell carcinoma (7 patients, 17 %) and large cell carcinoma (2 patients, 4.8%). Three individuals (7.3%) were diagnosed with non-small cell lung cancer without histological classification. Concerning stages, 5 cases were diagnosed as stage IA, 2 as stage IB, 2 as stage IIA, 2 as stage IIB, 12 as stage IIIA, 4 as stage IIIB and 7 as stage IV. Clinical differences among the different stages are specified in Supplementary Table 1. The remaining 7 individuals were diagnosed with small cell lung cancer at advanced stage. The reduced number of patients for each diagnostic and stage did not enable the inclusion of these variability sources in the study.

The control group included 55 healthy individuals, 24 of which were active smokers. The existence of lung cancer or other severe pulmonary diseases (interstitial pneumonitis, pneumonia, tuberculosis, etc.) was excluded by fibrobronchoscopy and/or computed tomography scanning. Exclusion criteria

were: (i) coexistence of extrapulmonary neoplasia during the last five years or chemotherapy for other different neoplasias; (ii) severe organ disease with negative influence on prognostic or with influence on the application of the target protocol, including congestive heart failure, chronic liver disease (functional stage MELD >12), chronic kidney disease stage 5 with substitutive treatment (hemodialysis or peritoneal dialysis). The control group had an average age of 55±10 years and 71% of them were males.

The study was approved by the ethical committee of the Reina Sofia University Hospital. The individuals selected for this study were previously informed to obtain complete consent. All steps from sweat extraction to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki of 2004.

It should be emphasized that sweat collection was performed before finishing the clinical study, which means that patients were not subjected to cytostatic treatment. Therefore, medication was exclusively restricted to control the symptoms by analgesics (paracetamol or ibuprofen), cough suppressants (codeine) and similar drugs. Apart from that, both control individuals and patients were taking common treatments related to diseases associated to the age interval (statins, antihypertensives, analgesics, etc.).

2.4. Sweat extraction

A Macroduct® SweatAnalysis System (Wescor, Utah, USA), consisting of a Webster sweat inducer and a Macroduct sweat collector (US Patent 4,542,751), was used. Pilogel® Iontophoretic Discs (US Patent 4,383,529) (Wescor, Utah, USA), a gel reservoir of pilocarpinium ions, were used in the iontophoretic stimulation of sweat excretion. A total sweat volume around 50 µL was collected per individual.

The sweat inducer provided a current intensity of 1.5 mA for 5 min through two Pilogel discs as electrodes located on the forearm. After removing the discs the skin where the positive disc had been located was cleaned with distilled water and the Macroduct collector covered this skin to collect sweat for

15 min. The collected sweat was transferred to a microEppendorf tube and stored at –80 °C until use.

2.5. Sample treatment

Sweat samples (10 μL) were diluted with 0.1% formic acid in 1:2 (v/v) ratio. The mixture was vortexed for 1 min and placed in the LC autosampler for subsequent analysis. All samples were prepared by duplicate.

A pool from each group under study (lung cancer and control patients) was prepared taking 5 μL aliquots from each participant. The two pools were 1:2 diluted with water containing 0.1% formic acid, and vortexed for homogeneization before analysis.

2.6. LC–QTOF MS/MS analysis

Chromatographic separation was performed using a C18 reverse-phase analytical column (Mediterranean, 50 mm × 0.46 mm i.d., 3 μm particle size) from Teknokroma (Barcelona, Spain), which was thermostated at 25°C. The mobile phases were water (phase A) and ACN (phase B) both with 0.1% formic acid as ionization agent. The LC pump was programmed with a flow rate of 0.8 mL/min with the following elution gradient: 3% phase B as initial mobile phase was kept constant from min 0 to 1; from 3 to 80% of phase B from min 1 to 10.5, and from 80 to 100% of phase B from min 10.5 to 11.5. A post-time of 5 min was set to equilibrate the initial conditions for the next analysis. The injection volume was 3 μL and the injector needle was washed for 10 times between injections with 70% methanol. Furthermore, the needle seat back was flushed for 15 s at a flow rate of 4 mL/min with 70% methanol to avoid cross contamination.

The parameters of the electrospray ionization source, operating in negative and positive ionization mode, were as follows: the capillary and fragmentor voltage were set at ±3.5 kV and 175 V, respectively; N_2 in the nebulizer was flowed at 40 psi; the flow rate and temperature of the N_2 as drying gas were 8 L/min and 350 °C, respectively. The instrument was calibrated and tuned according to procedures recommended by the manufacturer. MS and

MS/MS data were collected in both polarities using the centroid mode at a rate of 1 spectrum per second in the extended dynamic range mode (2 GHz). Accurate mass spectra in MS scan and MS/MS mode were acquired in the *m/z* range 60–1100. The instrument gave typical resolution 15000 FWHM (Full Width at Half Maximum) at *m/z* 118.086255 and 30000 FWHM at *m/z* 922.009798. To assure the desired mass accuracy of recorded ions, continuous internal calibration was performed during analyses by using the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated hexakis (1H, 1H, 3H-tetrafluoro-propoxy) phosphazine or HP-921] in positive ion mode; while in negative ion mode ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct of HP-921) were used.

Samples were first injected in full scan acquisition mode (two replicates per sample) and then in auto MS/MS acquisition mode with a preferred list to obtain the product ion information of metabolites with significant difference between the groups under study. The maximum number of precursors selected per cycle was set at 2, with an exclusion window of 0.3 min after 2 consecutive selections of the same precursor. Three collision energies (10, 20 or 40 eV) were used to obtain the maximum fragmentation information for each entity.

2.7. Data processing and statistical analysis

MassHunter Workstation software (version 3.01 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) was used to process all data obtained by LC–QTOF in full scan 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 exceeding 500 counts with a single charge state. Additionally, the isotopic distribution for a 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). Adducts formation in the positive (+H, +Na) and negative ionization (–H, +HCOO) modes as well as neutral loss by dehydration were also included to identify features corresponding to the same molecule. Thus, ions with identical elution profiles and related *m/z*

values (representing different adducts or isotopes of the same compound) were extracted as entities characterized by retention time (RT), intensity in the apex of chromatographic peaks and accurate mass. Background contribution was removed by subtraction of MFs linked to plasticizers, solvent impurities and other contaminants after analysis of a blank (0.1% formic acid) under identical instrument operational conditions. In this way, raw data files were created in compound exchange format (.cef files) for each sample and exported into the Mass Profiler Professional (MPP) software package (version 2.2, Agilent Technologies, Santa Clara, CA, USA) for further processing.

In a next step, the data were preprocessed by alignment of retention times and *m/z* values across the data matrix using a tolerance window of 0.4 min and 10 ppm mass accuracy, respectively. The alignment was based on the iterative comparison of *m/z* and retention times values for all MFs detected in each sample with the resting samples. Data pretreatment was based on baselining for each variable across samples to treat all compounds equally regardless of their intensity. This involved a preprocessing method that subtract the median abundance of each variable from the corresponding values in each sample. Furthermore, a logarithmic transformation was used to reduce relatively large differences among the respective MF abundances. Stepwise reduction of MFs number was performed based on frequency of occurrence and abundance of the respective MFs in classes. This filter allowed retaining MFs detected in all the samples from each group (cancer and control), removing those entities only detected in reduced number of samples, which is due to individual variability. MPP software also allowed unsupervised and supervised analysis of the data by PCA and PLS-DA, respectively. In both cases, the data set used was the previously scaled. An N-Fold cross-validation model was selected for this research. With this model, the individuals 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 repeated N times, with a different subset iteratively used for testing. Thus, each object 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 3 were selected in all PLS-DA models.

Identification of the most relevant entities was done using MS and MS/MS information and searching in the METLIN MS and MS/MS database (http://metlin.scripps.edu) and Human Metabolome Database (HMDB, 3.5 version). Statistical significance of metabolites identified was done by Kruskall-Wallis analysis.

The R (URL http://www.R-project.org) and pROC (1.5.1 version, URL http://web.expasy.org/pROC/) were used to evaluate the prediction performances of each identified compound. PanelomiX toolbox, supported on the iterative combination of biomarkers and thresholds, was used to combine biomarkers by selecting thresholds that provide optimal classification performance (Robin et al. 2013). Two panels of markers were built by setting either sensitivity or specificity cut-off values > 80%. In both cases, 5 repetitions and a fold number of 10 were selected. Cross-validation was executed using the PanelomiX computational toolbox, and no pre-filtering steps were included.

3. Results and discussion

3.1. Quality control and data pretreatment

The pool of sweat prepared from each group under study (lung cancer patients and control individuals) was injected 10 times in each polarity mode to obtain the global profile of molecular features associated to them. The initial data set was built after alignment of LC–TOF/MS chromatograms for efficient extraction of MFs according to retention time and mass accuracy. The tolerance window for both parameters was set at 0.4 min and 10 ppm for elution time and mass accuracy, respectively.

After data pretreatment, two data sets were obtained for positive and negative ionization modes. The experimental variability was minimized by applying a filter algorithm based on the detection frequency of MFs in the

replicates. In particular, the algorithm eliminates MFs no detected in at least all the replicates from each group —in this case, patients with lung cancer and control group. The application of the frequency filter simplified the data sets to 209 and 35 MFs in negative and positive ionization modes, respectively. These sets were used to carry out a quality control (QC) study to evaluate the instrumental precision of the detection step. With this aim, a PCA was built for data sets (Supplementary Fig. 1) that revealed the absence of variability sources due to time injection. Under these conditions, the relative standard deviation for MS signals was below 20% in both positive and negative ionization modes.

The two sets of MFs obtained in both polarities using the pools were extracted in data files obtained by analysis of all individual samples and aligned with the same tolerance parameters used for pools. As samples were injected by duplicate, the average area of the two injections was calculated for each potential molecular feature. Then, the frequency filter was applied to remove those MFs no detected in at least a representative percentage of samples forming each predefined class. This representative percentage was set at 75%, selection supported on statistical criteria to build a fingerprinting representative of the sweat metabolome from the two target groups. It is worth mentioning that the pathology influences differently the serum metabolome of the individuals. For this reason, the filtering algorithm should not be completely restrictive. After this filter, 33 and 28 potential molecular entities were found in the positive and negative ionization modes.

3.2. Identification and statistical analysis

Once the molecular entities with statistical significance were established, the MS/MS spectra of these entities were obtained by LC–QTOF MS/MS in automated mode to obtain the necessary information for identification. In this mode, the selection of precursor ions is automatically carried out in each MS scan, but a list of preferred ions can be included for selective fragmentation. The list of preferred precursor ions was formed by those associated to the molecular entities of the data set. The HMDB and METLIN database were used for

identification. To compare the MS/MS spectra of each entity with those recorded in the METLIN database, the information was collected using 10, 20 and 40 eV as collision energies with the purpose of obtaining different fragmentation levels. Table 1 shows the list of identified metabolites including the retention time and *m/z* values for precursor and product ions. Exogenous metabolites and interferents present in the Pilogel® to induce sweat production (*e.g.* pilocarpine or propylparaben) were removed from the data set. Most of the identified compounds were detected in negative ionization mode. As a result, only three metabolites (tryptophan, maltotetraose and 1-octen-3yl-glucoside) were confirmed in the positive ionization mode for inclusion in a preliminary panel.

Table 1. Metabolites identified in sweat, including the retention time and *m/z* values for precursor and product ions as well as the error in precursor ion detection expressed as ppm.

Negative ionization mode					
	Compound name	**RT (min)**	**Precursor ion (*m/z*)**	**Δppm**	**Product ions (*m/z*)**
1	Citrulline	0.7	174.0883	0	131.0818 44.9974 72.0078
2	Trisaccharide phosphate $C_{18}H_{34}O_{20}P$	0.8	601.1371		547.0835 96.9696 161.0440 259.0221
3	Phenylalanine	3.5	164.0718	0	147.0445 103.0549 91.0549 72.0094
4	Nonanedioic acid	7	187.0981	2	125.0962 187.0954 97.0645 123.0799 57.0335
5	Sulfonic lipid	8.5	577.1470		170.0046 391.1109 497.1911
6	Maltotriose	0.7	549.1672	0	503.1598 341.1085 179.0553 89.0241

7	Histidine	0.6	154.0624	1	137.0358 93.0457 80.0385 67.0307
8	Taurine	0.7	124.0072	1	79.9567
9	Urocanic acid	1.2	137.0360	2	93.0445 66.0349 65.0206 50.0050
10	Muconic acid	7.5	123.0089	5	53.0033 95.0139 68.9982
11	Suberic acid	6.2	173.0824	2	111.0809 83.0503 57.0349
12	MG(22:2)	8.7	431.3127	3.6	89.0247
Positive ionization mode					
	Compound name	**RT (min)**	**Precursor ion *(m/z)***	**Δppm**	**Product ions *(m/z)***
1	Tryptophan	4.7	205.0970	0.7	188.0699 146.0593 118.0647
2	Maltotetraose	0.8	689.2103	1	145.0488 127.0385 85.0283 487.1640
3	1-Octen-3yl-glucoside	8.4	291.1802	0.1	273.1697 127.1117 57.0899

After metabolites identification, quantitative information (height of the chromatographic peaks obtained from extracted ion chromatograms, EICs) was extracted in all samples in the studied cohort. Statistical analysis by the Kruskal-Wallis test between the two groups studied (lung cancer patients and control individuals) revealed that the trisaccharide phosphate, MG(22:2), nonanedioic acid, maltotetraose and maltotriose were the most significant metabolites (p-value < 0.008), as Table 2 shows. Urocanic acid and γ-GluLeu dipeptide were also significant presenting a p-value between 0.01 and 0.05. The remaining

metabolites present in the data set did not report significant difference between the two groups under study.

Table 2. Kruskal-Wallis statistical analysis to evaluate the significance of the identified metabolites in sweat to discriminate between lung cancer patients and control individuals.

Marker	p-Value
Trisaccharide phosphate	0.00019
Maltotriose	0.00177
Nonanedioic acid	0.00194
MG (22:2)	0.00411
Maltotetraose	0.00728
γ-GluLeu	0.03809
Urocanic acid	0.04437
Suberic acid	0.07485
Sulfonic lipid	0.1828
Citrulline	0.3254
Tryptophan	0.3417
Histidine	0.5015
Phenylalanine	0.5648
Muconic acid	0.6763
GluVal	0.7430
Taurine	0.8569

3.3. Multivariate analysis for discrimination of lung cancer patients and control individuals

It is worth mentioning that the objective of this research was not to establish metabolite differences among histological types as presently the cyto-histological diagnostic to confirm both tumour and hystological subtypes is irreplaceable. The same situation is found with the cancer stage. It would be very interesting to know if there are differences among the stages of the disease, which could be the subject of future research. Presently, distinction among stages requires x-ray computed tomography and endoscopic techniques to evaluate the endobronchial lesions and to study lymphadenopathies in the mediastinum.

As commented above, one of the problems concerning lung cancer diagnosis is the stage at which the disease is detected and the characteristics of the existing available tests (cost and complexity), which hinders their application to all patients. Therefore, more economical tests with prognostic capability are demanded for a preliminary screening, which allows reducing the population under study by eliminating individuals in a categorical "rule-out" group. For this reason, the most important aspect in a prediction model for lung cancer is a low probability of false negatives, while the probability of false positives is not critical as the disease could be confirmed by other tests.

In this study a discrimination model based on PLS-DA was built using the dataset formed by identified metabolites (Table 1). The accuracy parameters for the validation and training sets of the resulting PLS-DA are listed in Table 3,while the scores plot is shown in Fig. 1. As Table 3 shows, according to the training set the probability of false positives was 42.7%, which is high. However, the presence of the pathology for these target individuals could be confirmed with one other test. On the other hand, the probability of false negatives was only 13.3%. Therefore, this discrimination model led to know that the identified metabolites could be combined to reduce the proportion of individuals demanding for a confirmatory test.

Table 3. Accuracy parameters for the validation and training sets of the developed PLS-DA model including sensitivity and specificity and the positive and negative predictive value.

Parameters / Set	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
Training	79.8	93.1	90.5	84.7
Validation	78.6	86.1	82.5	82.8

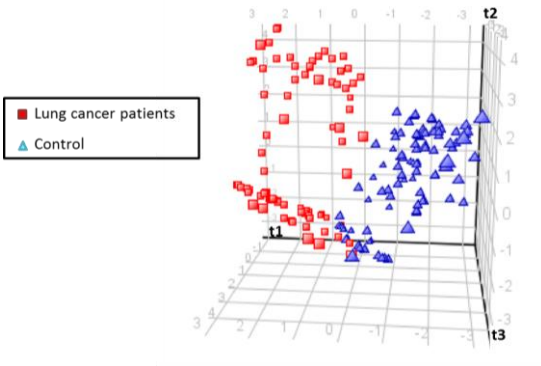


Fig. 1. PLS-DA scores plot for discrimination of lung cancer patients and control individuals by sweat metabolomics analysis

3.4. Statistical evaluation of the prediction capability by ROC curves

Assessment of sensitivity and specificity of the significant metabolites for lung cancer prediction was carried out by obtaining the ROC curves and estimating the partial area under the curve (pAUC). Evaluation of a predictor by the total AUC is not recommended when the performance test only takes place in high specificity or high sensitivity regions (Robin et al. 2011). For this reason, the pAUC parameter is more suited since it is restricted to specific regions of the curve. In this case the region selected encompassed from 80 to 100 % of specificity to reduce the probability of false negative predictions. With this restriction, the maximum value of pAUC obtained would be 20%. The values of pAUC, sensitivity and specificity expressed in percentage, with a 95% of confidence interval, are listed in Table 4, and ROC curves obtained for metabolites with the highest discrimination capability are shown in Figure 2. As can be seen, the compounds that presented a higher value of pAUC were the trisaccharide phosphate (10.6%), maltotriose (10.0%), nonanedioic acid (7.9%) and maltotetraose (6.4%). These data show that the marker with the best discrimination capability of this dataset was the trisaccharide phosphate.

Regarding specificity and sensitivity, the compound with the highest specificity was muconic acid (100%), followed by maltotetraose, the sulfonic

lipid, GluVal dipeptide and histidine, all with 97.1% of specificity for discrimination of patients with lung cancer from control individuals. However, any of these compounds exhibits the highest sensitivity, which was attributed to the trisaccharide phosphate (72.7%). This fact supports the necessity for combining metabolites to develop a panel of markers in order to obtain statistical models with improved prediction capability.

Table 4. Parameters of the ROC curves obtained for each identified metabolite by selecting a specificity range from 80 to 100%. pAUC, sensitivity and specificity are included with a 95% of confidence interval (CI).

Marker	**% pAUC (95% CI)**	**% Specificity (95% CI)**	**% Sensitivity (95% CI)**
Trisaccharide phosphate	10.6 (6.5–15.3)	80.0 (65.7–91.5)	72.7 (54.5–90.9)
Maltotriose	10.0 (5.2–14.9)	91.4 (80.0–100.0)	63.6 (45.5–81.8)
Nonanedioic acid	7.9 (4.0–12.6)	88.6 (77.1–97.1)	50.0 (31.8–68.2)
Urocanic acid	6.6 (2.6–11.8)	94.3 (85.7–100.0)	40.9 (22.7–63.6)
Maltotetraose	6.4 (2.7–10.8)	97.1 (91.4–100.0)	31.8 (13.6–50.0)
Sulfonic lipid	5.5 (2.1–9.7)	97.1 (91.4–100.0)	27.3 (9.1–45.5)
MG (22:2)	4.8 (1.7–9.9)	94.3 (85.7–100.0)	31.8 (13.6–54.5)
γ-GluLeu	4.0 (0.5–9.1)	82.9 (68.6–94.3)	45.5 (22.7–68.2)
Muconic acid	3.0 (0.9–7.1)	100.0 (100.0–100.0)	13.6 (0.0–27.3)
Citrulline	3.0 (0.4–6.8)	82.9 (68.6–94.3)	31.8 (13.6–54.5)
Phenylalanine	2.7 (0.7–6.3)	82.9 (68.6–94.3)	27.3 (9.1–45.5)
Suberic acid	2.5 (0.4–8.0)	80.0 (65.7–91.4)	31.8 (13.6–54.5)
Tryptophan	2.3 (0.1–5.8)	94.3 (85.7–100.0)	13.6 (0.0–31.8)
GluVal	1.6 (0.0–5.7)	97.1(91.4–100.0)	4.5 (0.0–13.6)
Histidine	1.6 (0.0–4.8)	97.1 (91.4–100.0)	9.1 (0.0–22.7)
Taurine	1.2 (0.0–4.4)	82.9 (68.6–94.3)	18.2 (4.5–36.4)

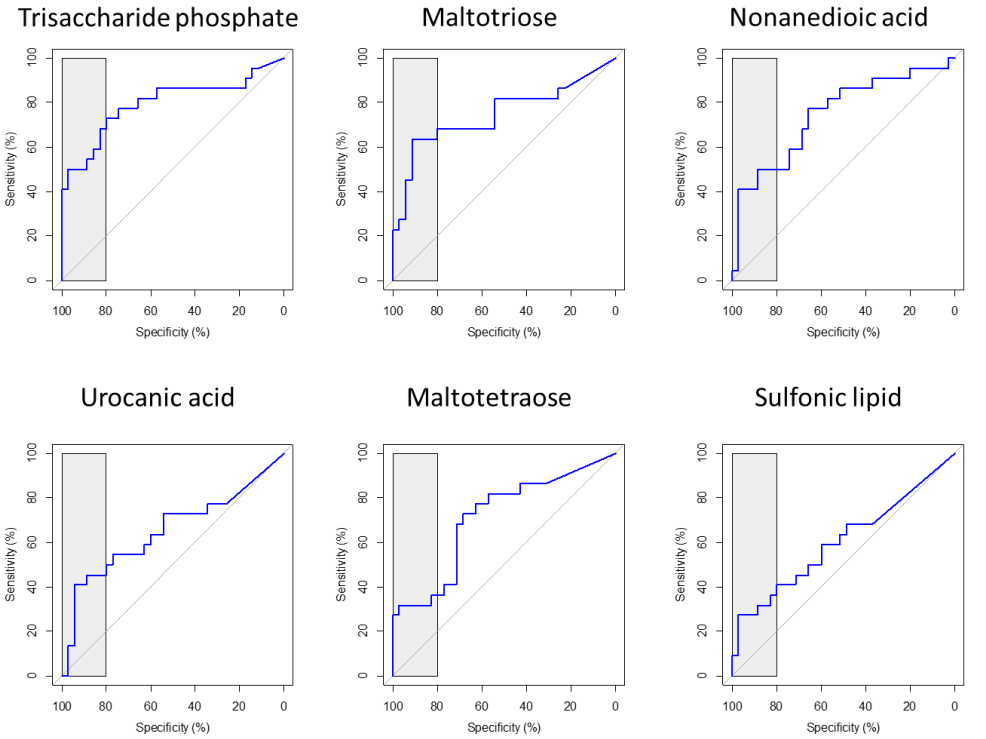


Fig. 2. ROC curves obtained for metabolites with the best prediction capability to discriminate lung cancer patients and control individuals. pAUC values calculated for specificity above 80% are included.

3.5. Development of panels with predction capability for lung cancer

The development of panels for discrimination between lung cancer patients and control individuals selected for this research was addressed by using PanelomiX as computational toolbox. This tool iteratively combines significant variables and thresholds to find panels of potential biomarkers with good prediction/discrimination parameters in terms of sensitivity and specificity. As mentioned above, the most important aspect in a prediction model for lung cancer is to reduce the probability of false negatives, while the probability of false positives is not critical as the disease could be confirmed by other tests.

Nevertheless, the development of discrimination models targeted at reducing false positives would also be useful as screening strategy to minimize the number of patients subjected to invasive confirmatory tests. In this research, the configuration of panels was supported on two statistical criteria: at least 80% sensitivity or specificity, which resulted in two panels of biomarkers. The panel with at least 80% specificity should allow decreasing considerably false negative rates. On the other hand, a biomarkers panel with at least 80% sensitivity should allow a significant decrease of false positive rates. PanelomiX was applied to the whole cohort to define panels of three biomarkers with the best classification parameters. The list of all identified metabolites (Table 1), in both positive and negative ionization modes, was considered as the set of variables. Cross-validation was included in the development of both panels for internal validation, which reported information on the model stability. The classification capability of the panels was assessed by threshold sensitivity and specificity and the pAUC value. The best panel providing specificity above 80% was obtained with the combination of three metabolites including nonanedioic acid, γ-GluLeu and maltotriose (100.0% specificity and 63.6% sensitivity, while the pAUC was 13.3% (95% CI 9.6–16.9). Cross-validation for this model also reported good statistics parameters since specificity and sensitivity were 80.0% and 59.1%, respectively. This panel would allow diagnosing individuals with lung cancer with a low false negative rate. Figure 3.A shows the ROC curve obtained with this panel in comparison with that obtained independently for trisaccharide phosphate, which was the best independent marker of this configuration in terms of pAUC value. As can be seen, sensitivity was significantly improved by application of the panel.

On the other hand, the best panel providing sensitivity above 80% was obtained with two of the metabolites included in the previous panel (nonanedioic acid and maltotriose) and MG(22:2). Figure 3.B shows the ROC curve from this second panel together with that reported by the best independent marker of this configuration, which was MG(22:2). The parameters obtained for this panel were 88.6% of specificity, 81.8% of sensitivity and a pAUC of 10.7% (95% CI 6.2–16.6).

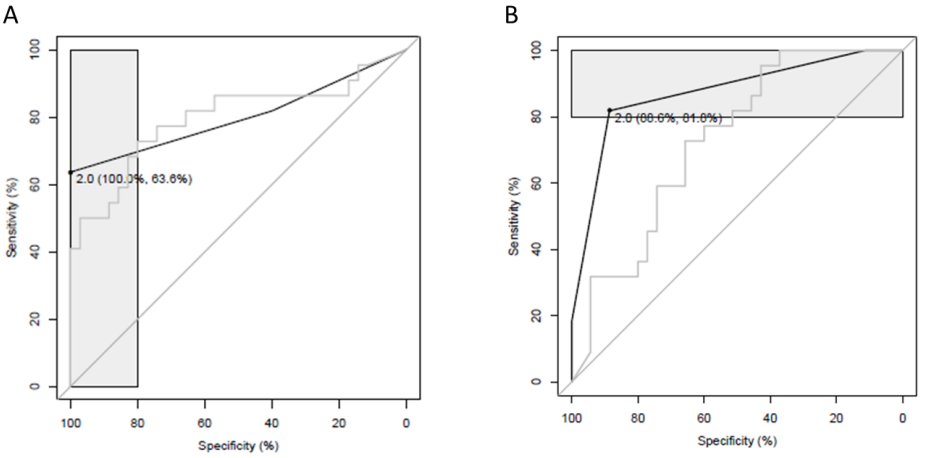


Fig. 3. ROC curves obtained from the configured panels with: (**A)** at least 80% specificity to reduce false negatives and, (**B)** at least 80% sensitivity to reduce false positives. The ROC curves are compared with those provided by the marker with the best independent prediction capability, trisaccharide phosphate and MG(22:0), which are shown in grey.

As both panels were characterized by high sensitivity and specificity values, a proposed approach could be supported on the analysis of the four involved metabolites. The combined use of the two panels would allow reducing the number of cases subjected to confirmatory test with the minimum rate of false negative and positive rates.

4. Conclusions

As previously exposed, one of the main limitations relative to the diagnostic of lung cancer is the stage at which this is detected as well as the characteristics of the existing available tests. These are invasive and require a high cost, which makes their application to all individuals defined by risk factors of lung cancer difficult. For this reason, more economical tests featured with high prediction capability are demanded for a preliminary screening, which should allow decreasing the number of individuals requiring confirmatory tests.

Sweat has not been widely exploited in the clinical field despite the advantages associated to its sampling. The potential of sweat as biofluid for implementation in the diagnostic of lung cancer has been demonstrated. A prediction model based on a panel of metabolites including amino acids, sugars and some lipids has been built to discriminate patients with lung cancer from a control group. The high negative predictive value reflects the potential use of this biofluid to reduce the number of cases as possible positive cases to be subjected to confirmatory tests.

The discrimination capability of individual metabolites revealed that a trisaccharide phosphate presented the highest prediction capability. However, this compound is not present in the two panels of three metabolites generated for lung cancer prediction. The two panels included maltotriose and nonanedioic acid in combination with γ-GluLeu and MG(22:2). Both panels significantly improved the discrimination capability of independent metabolites. The first panel was characterized by 100% specificity and 63.6% sensitivity. Therefore, the false negative rate was 0%. The second panel reported specificity and sensitivity values above 81%.

These preliminary results emphasized the necessity of a large scale study to validate the proposed panels in order to reduce the proportion of individuals to be subjected to confirmatory tests, and to detect lung cancer in less advances stages.

Acknowledgements

The Spanish Ministerio de Economía y Competitividad (MINECO) and the FEDER Program (project No CTQ2012-37428), and the Junta de Andalucía through an Excellence project (FQM-1602) are gratefully acknowledged for financial support. F. Priego-Capote is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks the MICINN for an FPU scholarship (AP2009-

0499). The Reina Sofía Hospital (Córdoba, Spain) is also grateful to donors of sweat samples.

References

Álvarez-Sánchez, B., Priego-Capote, F., & Luque de Castro, M.D. (2010). Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. Trends in Analytical Chemistry, 29(2), 111–119.

Beane, J., Vick, J., Schembri, F., et al. (2011). Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. Cancer Prevention Research, 4, 803–817.

Beger, R.D. (2013). A review of applications of Metabolomics in cancer. Metabolites, 3, 552–574.

Carrola, J., Rocha, C.M., Barros, A.S., Gil, A.M., Goodfellow, B.J., Carreira, I.M., Bernardo, J., Gomes, A., Sousa, V., Carvalho, L., & Duarte, L.F. (2011). Metabolic signatures of lung cancer in biofluids: NMR-based Metabonomics of urine. Journal of Proteome Research, 10(1), 221–230.

De Giovanni, & N., Fucci, N. (2013). The current status of sweat testing for drugs of abuse: a review. Current Medicinal Chemistry, 20(4) 545–561.

Fan, T.W.M., Lane, A.N., Higashi, R.M., et al. (2009). Altered regulation of metabolic pathways in human lung cancer discerned by 13C stable isotope-resolved metabolomics (SIRM). Molecular Cancer, 8(41), doi:10.1186/1476-4598-8-41.

Gibson, L.E., & Cooke, R.E. (1959). A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. Pediatrics, 23(3) 545–549.

Girard, L., Zöchbauer-Müller, S., Virmani, A.K., Gazdar, A.F., & Minna, J.D. (2000). Genome-wide allelotyping of lung cancer identifies new regions of allelic

loss, differences between small cell lung cancer and non-small cell lung cancer and loci clustering. Cancer Research, 60, 4894–4906.

Hew-Butler, T., Noakes, T.D., Soldin, S.J., & Verbalis, J.G. (2010). Acute changes in arginine vasopressin, sweat, urine and serum sodium concentrations in exercising humans: does a coordinated homeostatic relationship exist? British Journal of Sports Medicine, 44, 710–715.

Howlader, N., Noone, A.M., Krapcho, M., et al. (2013). SEER Cancer Statistics Review 1975–2010, National Cancer Institute.

Jentzmik, F., Stephan, C., Miller, K., Schrader, M., Erbersdobler, A., Kristiansen, G., Lein, M., & Jung, K. (2010). Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. European Urology, 58(1), 12–18.

Kami, K., Fujimori, T., Sato, H., et al. (2013). Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. Metabolomics, 9(2), 444–453.

Karlsson, M., Mårild, S., Brandberg, J., Lönn, L., Friberg, P., & Strandvik, B. (2006). Serum phospholipid fatty acids, adipose tissue, and metabolic markers in obese adolescents. Obesity, 14(11), 1931–1939.

Khadir, A., & Tiss, A. (2013). Proteomics approaches towards early detection and diagnosis of cancer. Journal of Carcinogene Mutagene, S14 002.

Kintz, P., Tracqui, A., & Mangin, P. (1996). Sweat testing in opioid users with a sweat patch. Journal of Analytical Toxicology, 20, 393–397.

Kirk, J.M., Keston, M., McIntosh, I., & Essa, S. (1992). Variation of sweat sodium and chloride with age in cystic fibrosis and normal populations: further investigations in equivocal cases. Annals of Clinical Biochemistry, 29 145–152.

Kutyshenko, V.P., Molchanov, M., Beskkaravayny, P., Uversky, V.N., & Timchenko, M.A. (2011). Analyzing and mapping sweat metabolomics by high-resolution NMR spectroscopy. PLoS one, 6(12) e28824.

Lee, N.V.L., Miller, P.W., & Buono, M.J. (2010). The effect of spironolactone on sweat and urinary sodium excretion during exercise in humans. Clinical Physiology and Functional Imaging, 30, 13–16.

Lynch, A. (2010) Thesis: Development and characterisation of solid-state ion-selective electrodes, and their application to sweat analysis for cystic fibrosis diagnosis, School of Chemical Sciences, Dublin City University.

Mena-Bravo, A., & Luque de Castro, M.D. (2014). Sweat: a sample with limited present applications and promising future in metabolomics. Journal of Pharmaceutical and Biomedical Analysis, doi:10.1016/j.jpba.2013.10.048.

National Lung Screening Trial Research Team. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England Journal of Medicine, 365(5), 395–409.

Pleasance, E.D., Stephens, P.J., O'Meara, S., et al. (2010). A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature, 463, 184–190.

Robin, X., Turck, N., Hainard, A., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77–84.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., & Müller, M. (2013). PanelomiX: a threshold-based algorithm to create panels of biomarkers. Translational Proteomics, 1(1), 57–64.

Sánchez de Cos, J., Hernández, J., Jiménez López, M., Padrones Sánchez, S., Rosell Gratacós, A., & Rami Porta, (2011). R. SEPAR Guidelines for lung cancer staging. Archivos de Bronconeumología, 47, 454–465.

Sears, M.E., Kerr, K.J., & Bray, R.I. (2012). Arsenic, cadmium, lead, and mercury in sweat: a systematic review. Journal of Environmental Public Health, article ID 184745.

Taylor, J.R. Watson, I.D., Tames, F.J., & Lowe, D. (1998). Detection of drug use in a methadone maintenance clinic: sweat patches versus urine testing. Addiction, 93 847–853.

Uemura, N., Nath, R.P., Harkey, M.R., Henderson, G.L., Mendelson, J., & Jones, R.T. (2004). Cocaine levels in sweat collection patches vary by location of patch placement and decline over time. Journal of. Analytical Toxicology, 28, 253–259.

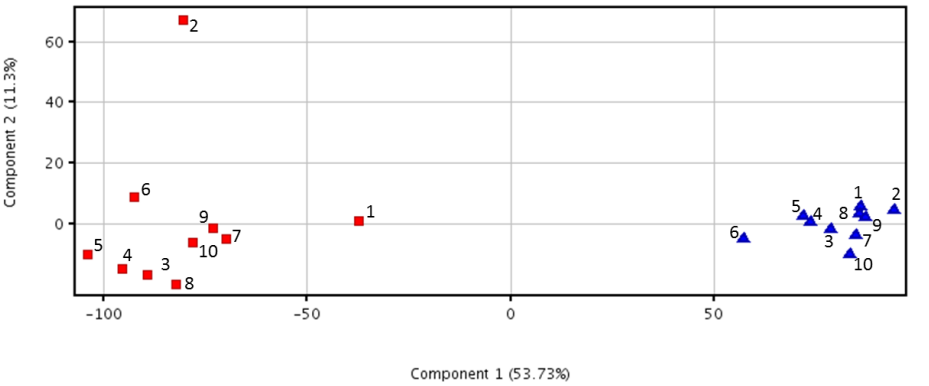
Wegwarth, O., Schwartz, L.M., Woloshin, S., Gaissmaier, W., & Gigerenzer, G. (2012). Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. Annals of Internal Medicine, 156(5), 340–349.

Wing, D., Prausnitz, M.R., & Buono, M.J. (2013). Skin pretreatment with micro needles prior to pilocarpine iontophoresis increases sweat production. Clinical Physiology and Functional Imaging, 33(6), 436–440.

World Health Organization. (2012). Cancer fact sheet No. 297. http://www.whoint/mediacentre/factsheets/fs297/en/_Accessed October 26, 2012.

Xiao, H., Zhang, L., Zhou, H., Lee, J.M., Garon, E.B., & Wong, D.T. (2012). Proteomic analysis of human saliva from lung cancer patients using two-dimensional difference gel electrophoresis and mass spectrometry. Molecular & Cellular Proteomics, 11(2) 1–12.

Supplementary material



Supplementary Fig. 1. Principal Component Analysis (PCA) built from the data set obtained after analysis (10 replicates) of sweat pools prepared from lung cancer patients and control individuals.

Supplementary Table 1. Main characteristics of the cohort under study including the age, sex and body mass index (BMI). All data are expressed as percent or mean (range).

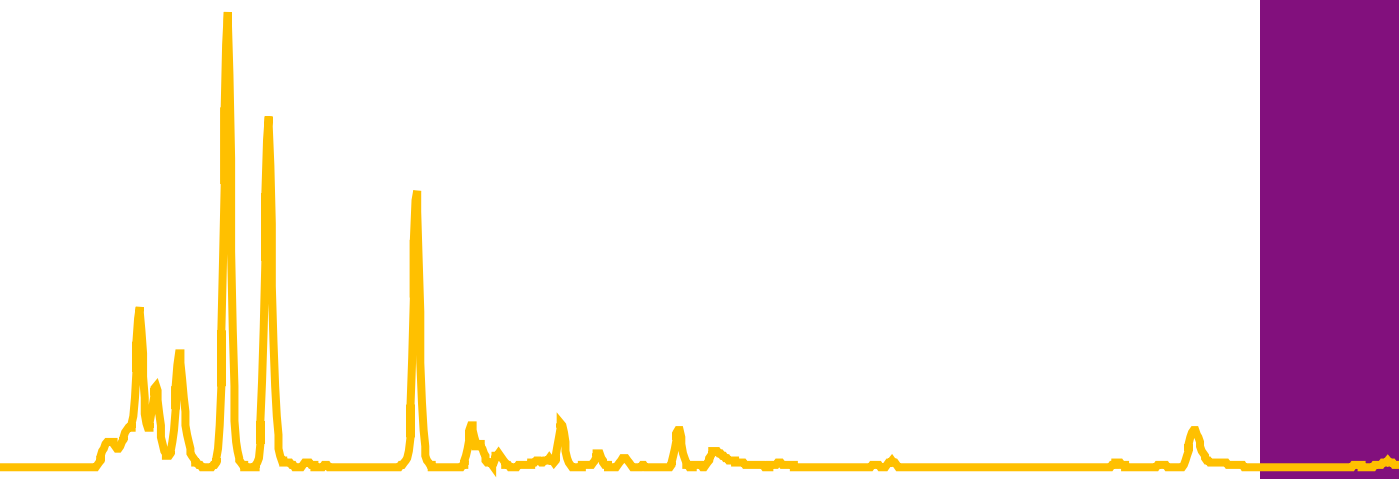
Individuals (n = 96)	**Lung cancer patients (n=41)**	**Controls (n=55)**		***P*-value[2]**
		Smokers (n=24)	**Non-smokers (n=31)**	
Age, *(year)*	63 (42–87)	56 (38–85)	54 (40–83)	0.344
Sex, *(male %)*	78	70	70	0.739
Pack-years[1]	51 (15–102)	37 (10–80)	0	
BMI, *(kg/m²)*	26.8 (17–38)	27.8 (22.6–36.3)	27.6 (21.8–40.1)	0.440

[1] Cigarettes consumption is included as pack-years. Pack-years = (average cigarettes smoked per day/20) x (years of smoking).

[2]P-values were obtained by analysis of variance (ANOVA) —the non-parametric version (Kruskal-Wallis) in the case of age as it does not fit normal distribution.

SECTION III

Targeted analysis of potential biomarkers



After development of approaches based on global profiling analysis in clinical metabolomics, the next step was to develop methods based on targeted or quantitative analysis of metabolites involved in relevant biological processes. All the protocols were developed by using human serum as biofluid.

Three of the targeted platforms have been applied to the analysis of different groups of metabolites in atherosclerotic patients. Thus, phospholipid profiles have been analyzed by LC–MS/MS in high resolution mode for evaluation of atherosclerotic patients with different diagnostics, thus leading to the proposal of a four-marker panel with prediction capability for stable angina and NSTEMI patients, as shows Chapter 8.

A hyphenated SPE–LC–MS/MS configuration for determination of eight essential amino acids allows discrimination between atherosclerotic patients with stable angina and acute myocardial infarction. The information, in Chapter 9, was supported on a cohort of 122 atherosclerosis patients, diagnosed with the aid of a catheterization test.

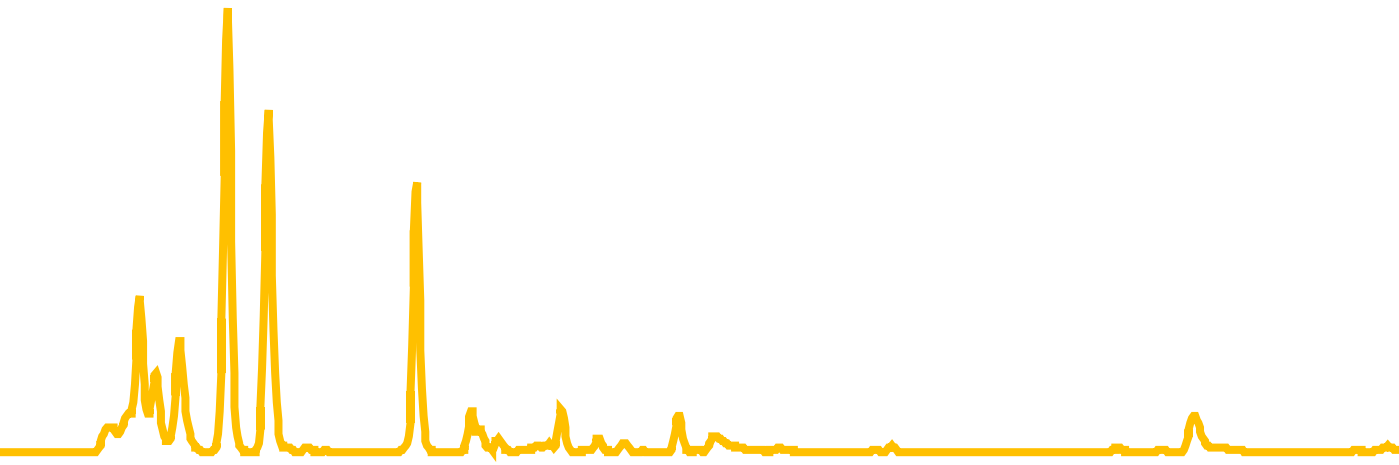
The content of Chapter 10 deals with metabolites involved in the tricarboxylic acids (TCA) cycle that had been previously proposed as cardiovascular biomarkers. A method for determination of the target metabolites based on GC–MS applied to serum from 223 patients —172 of them with significant coronary lesion— has demonstrated the influence of three cardiovascular risk factors such as obesity, hypercholesterolemia, and smoking habit on serum levels of TCA-cycle metabolites. Multifactor analysis of variance and statistical evaluation by ROC curves support interactions between the occurrence of a coronary lesion and the risk factors considered in this study, and the discrimination capability of these metabolites, respectively.

Assessment of cathelicidin as a *biomarker* of prone to infectious diseases status made mandatory a robust, automated, selective and sensitive platform for

determination of the peptide in serum samples. Chapter 11 shows such platform in which robustness was achieved by unattended operation of the approach, automation was allowed by online coupling of a commercial system for SPE with a triple quadrupole mass spectrometer, and selectivity and sensitivity were achieved by the joint effect of removal of interferents non-retained on the sorbent, and preconcentration in the solid-phase cartridge and the MS detector, respectively. In this way, short analysis time, and detection and quan-titation limits at the low μg/L levels were achieved.

Chapter 8:

Analysis of serum phospholipid profiles by LC–MS/MS in high resolution mode for evaluation of atherosclerotic patients



Accepted in
Journal of Chromatography A

Analysis of serum phospholipid profiles by LC–MS/MS in high resolution mode for evaluation of atherosclerotic patients

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b]*, *J.G. Galache-Osuna*[c], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Department of Cardiology, Hemodynamic and Interventional Cardiology, University Hospital Miguel Servet, Zaragoza, Spain.*

Analysis of serum phospholipid profiles by LC–MS/MS in high resolution mode for evaluation of atherosclerotic patients

*M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M. D. Luque de Castro**

Abstract

Atherosclerosis is one of the most frequent aetiology for myocardial infarction and death. It is also the main cause of angina pectoris, a perceived symptom that results of a mismatch between myocardial supply and demand caused by the partial obstruction of the arteries. The correct diagnosis of atherosclerosis can led to a reduction of mortality. In this research, the phospholipids profile of serum samples from patients diagnosed with stable angina and non-ST elevation myocardial infarction (NSTEMI) or unstable angina have been compared to found potential markers to discriminate between stable angina and NSTEMI. Phospholipids profiles obtained by LC–QTOF MS/MS have led to the identification of four potential markers with significant difference between the two groups under consideration: lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0). The panel generated with the four compounds have a disease prediction capability in the training set of 70.7 and 66.0% for stable angina and NSTEMI patients, respectively. On the other hand, the validation set by application to an independent cohort improved the predictive power for angina stable patients (92.3%) while this was slightly decreased up to 50.0% for NSTEMI patients. The discrimination capability of the metabolite panel was assessed in terms of sensitivity and specificity by ROC analysis. In this way, LC–QTOF MS/MS has shown to be a useful strategy for phospholipid profiling in serum and development of tools to aid in clinical diagnostic.

1. Introduction

Among of the most frequent causes of death worldwide are cardiovascular diseases and, in particular, atherosclerosis, with over 80% of deaths in low- and middle-income countries [1]. Atherosclerosis is the main cause of angina pectoris, a perceived symptom that results from a mismatch between myocardial supply and the demand caused by partial obstruction of the arteries. When the anaerobic threshold is crossed, the patient develops symptomatic angina pectoris (stable angina) that can appear after a physical effort or activity [2]. If this symptom is perceived with minimum activity and high frequency, the suited term is unstable angina. A rupture in the atherosclerotic plaque can generate a thrombus that can occlude partially or completely the normal blood flow. Complete thrombotic occlusion of a coronary artery causes myocardial necrosis and is termed acute ST-elevation myocardial infarction, also known as acute myocardial infarction (AMI), whereas incomplete occlusion does not cause myocardial necrosis and is termed non-ST elevation myocardial infarction (NSTEMI). Unstable angina and NSTEMI are often considered together because they can be indistinguishable upon first appearance [3].

One of the main problems in clinical practice is the diagnosis of cardiovascular dysfunctions that are commonly detected in a relatively advanced stage of the disease course [4]. The initial diagnostic approach for coronary artery disease encompasses a detailed patient history, a complete physical examination, and an electrocardiogram, which is one of the test that supports the discrimination between stable angina and NSTEMI. The main limitation of the electrocardiogram is that its profile could be also altered by other cardiopathies such as acute pericarditis or left ventricular hypertrophy, resulting in false positive results [5]. After this preliminary evaluation, laboratory blood tests (troponin, creatine kinase,...), stress testing, and invasive cardiac catheterization may be necessary to obtain further insight of the diagnostic.

Blood lipoproteins contain a protein and a phospholipid component, which allow setting differences among them [6]. Extensive epidemiological

studies have revealed a major role of lipoproteins in the development of atherosclerosis, with special emphasis on LDL cholesterol, the concentration of which is directly related to the risk of coronary artery disease, while HDL cholesterol level is inversely related to this risk [7]. Despite LDL-lowering therapies are frequently recommended, cardiovascular events still occur. For this reason, the lipoprotein composition should play a relevant role in the occurrence and progression of atherosclerosis. In this sense, the plasma phospholipid transfer protein (PLTP) mediates the transfer of phospholipids from VLDL and LDL to HDL and it is able to modulate HDL size and composition [8]. Additionally, PLTP may also be involved in HDL cellular-mediated efflux of phospholipids and cholesterol [9]. The alteration of the content of negatively charged phospholipids in vascular cell membranes was also shown in patients with essential hypertension [10]. Other studies have suggested that the phospholipid composition of HDL could be a key factor in processes such as inhibition of adhesion molecules and also in its anti-inflammatory and anti-oxidant properties [11]. On the other hand, the phospholipid content is one of the factors contributing to the closer lateral molecular packing in LDL [12].

The objective of the present research was to compare the profile of phospholipids in serum from patients diagnosed with atherosclerosis after catheterization. Patients were classified in two different clinical groups: those with non-ST elevation myocardial infarction (NSTEMI) and those diagnosed with stable angina. Discrimination between both groups by multivariate methods are used to search for a predictive panel formed by marker phospholipids with capability to differentiate the two atherosclerotic diagnostics. Phospholipids analysis was accomplished by a selective solid-phase extraction (SPE) protocol with deproteinization and subsequent determination by liquid chromatography high resolution tandem mass spectrometry using a QTOF detector to take benefits from high accuracy for phospholipids identification.

2. Material and methods

2.1. Reagents

MS-grade formic acid and acetonitrile (ACN) from Scharlab (Barcelona, Spain) were used for SPE sample preparation and to prepare the chromatographic mobile phases. LC-grade ammonia and methanol were also from Scharlab. Deionized water (18 mΩ·cm) from a Millipore Milli-Q water purification system was used.

2.2. Instruments and apparatus

A vacuum manifold (Varian VacElut SPS 24, USA) was used for SPE–deproteinization of phospholipids from human serum. A speed-vac ConcentratorPlus, from Eppendorf Ibérica (Madrid, Spain), was used to evaporate the methanol phase after SPE elution to concentrate the analytical sample.

An Agilent 1200 Series LC system (consisting of a binary pump, a vacuum degasser, an autosampler and a thermostated column compartment) coupled to an Agilent 6540 UHD Accurate-Mass Q–TOF hybrid mass spectrometer equipped with a dual electrospray (ESI) source (Santa Clara, CA, USA) was used. The chromatographic eluate was monitored in high-resolution mode.

2.3. Cohort selected for the study

One hundred forty patients diagnosed with coronary atherosclerosis by catheterization were selected in the Cardiology Unit of Miguel Servet Hospital (Zaragoza, Spain). Two clinical diagnostics were distinguished in the selected cohort: unstable angina or non-ST elevation myocardial infarction (NSTEMI) diagnosed to 72 patients and stable angina to 68 patients. The main characteristics of the individuals forming the cohort under study are specified in Table 1. The independence of clinical factors and cardiovascular pathologies was proved according to the Pearson's Chi squared test for categorical factors and by the Mann-Whitney test for numerical factors with a 95% of confidence level in both cases. The *p-value* obtained for each factor is also listed in Table 1.

Table 1. Main characteristics of the cohort as well as the *p*-value obtained after studying the independence of each factor from the disease under study. The test used for categorical factors was the Pearson's Chi Squared test, while the Mann-Whitney test was used for numerical factors.

Factor	**Population (n=140)**	**Stable angina (n=68)**	**NSTEMI (n=72)**	***p-Value***
Age, *yr*	68 (33–96)	67 (42–82)	69 (33–96)	0.07
Sex, *% male*	74.9	79.4	70.6	0.08
Obesity, *% obese (BMI≥30)*	39.8	44.1	35.8	0.43
Hypercholesterolemia, *%*	54.4	54.4	54.4	0.66
Arterial hypertension, *%*	58.9	60.6	57.3	0.66
Renal insufficiency, *%*	6.7	2.9	10.4	0.11
Diabetes, *%*	24.2	26.5	22.1	0.58
Smoking habit, *%*	49.8	54.4	45.6	0.45

Most atherosclerotic patients were taken medication as follows: β-blocking agents (78%), statins (92%), antiplatelet agents (88%) and aspirin (98%). However, distribution of patients under pharmaceutical treatment was equal for the two groups and, statistical analysis did not show significance for any treatment.

2.4. Blood extraction and serum isolation

Blood extraction immediately preceded coronary catheterization. All steps from blood extraction to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004), which were supervised by specialized personnel from Miguel Servet Hospital that approved the experiments. Individuals selected for this study were informed to obtain consent prior to sample collection.

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton, Dickinson, Franklin Lakes, NJ, USA) and incubated for 30 min at room temperature to allow coagulation. Then, the tubes were centrifuged at 2000 *g* for 15 min at 4 °C to isolate the serum fraction. Serum was placed in plastic ware tubes and stored at −80 °C until analysis.

2.5. Sample treatment

Phospolipids were extracted from serum samples using 30 mg HybridSPE® cartridges from Supelco (PA, USA) with the following protocol. A 300 µL aliquot of serum of each patient was mixed with 700 µL of acetonitrile acidified with 1% formic acid (v/v) in the cartridge and left for 2 min. Then, the vacuum was applied to make the mixture to flow through the sorbent cartridge, which was then washed with 1 mL of acetonitrile acidified with 1% formic acid (v/v) and 1 mL of acetonitrile. Phospholipids were eluted by pH change using 1 mL of methanol with 5% (v/v) ammonium hydroxide.

The extract was evaporated and the resulting residue was reconstituted with 300 µL of methanol with 5% (v/v) ammonium hydroxide and shaken for 30 s before injection into the LC–QTOF MS/MS equipment.

2.6. LC–QTOF MS/MS analysis

Chromatographic separation was performed using a C18 reverse-phase analytical column (Mediterranean, 50 mm x 0.46 mm i.d., 3 µm particle size) from Teknokroma (Barcelona, Spain), which was thermostated at 25 °C. The mobile phases were water (phase A) and acetonitrile (phase B) both with 0.1% formic acid as ionization agent. The LC pump was programmed with a flow rate of 0.7 mL/min with the following elution gradient: 25% phase B was kept as initial mobile phase constant from min 0 to 2; from 25 to 100% of phase B from min 2 to 12, and kept at 100% for 3 min. A post-time of 5 min was set to equilibrate the initial conditions for the next analysis. The injection volume was 5 µL and the injector needle was washed between injections with 80% methanol for 10 times. Furthermore, the needle seat back was flushed for 12 s at a flow rate of 4 mL/min with 80% methanol to avoid cross contamination.

The parameters of the electrospray ionization source operating sequentially in negative and positive ionization modes were as follows: the capillary and fragmentor voltage were set at ±3.5 kV and 175 V, respectively; N_2 in the nebulizer was flowed at 60 psi; the flow rate and temperature of the N_2 as drying gas were 10 L/min and 350 °C, respectively. The instrument was

calibrated and tuned according to procedures recommended by the manufacturer. MS and MS/MS data were collected in both polarities using the centroid mode at a rate of 2.6 spectrum/s in the extended dynamic range mode (2Ghz). Accurate mass spectra in auto MS/MS mode were acquired in MS *m/z* range 60–1100 and MS/MS *m/z* range 60–1100. The instrument gave typical resolution 15000 FWHM (Full Width at Half Maximum) at *m/z* 118.086255 and 30000 FWHM at *m/z* 922.009798. To assure the desired mass accuracy of recorded ions, continuous internal calibration was performed during analyses by using the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonated hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazine or HP-921] in positive ion mode; while in negative ion mode ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formateadduct 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 25 V.

2.7. Data processing and statistical analysis

MassHunter Workstation software (vesion 5.00 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) was used to process all analysis data obtained by LC–QTOF in MS/MS mode. Identification of the different phospholipid families was carried out with a search algorithm including characteristic product ions and neutral losses of each phospholipid class and ions ascribed to the alkanoyl chains. Then, 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).

The dataset obtained after identification included the list of all phospholipids detected in serum samples from atherosclerotic patients of the cohort. This dataset was exported as compound exchange format (.cef) files for each sample and imported into the Mass Profiler Professional (MPP) software package (version 2.2, Agilent Technologies, Santa Clara, CA, USA) for further

processing. Preprocessing of data was based on the alignment of compounds and logarithmic transformation to reduce relatively large differences among the respective abundances of phospholipids in serum.

MPP software allowed statistical analysis of data to detect variability sources as well as supervised analysis of the data by Partial Least Squares–Data Analysis (PLS–DA). The autoscaling was set as pretreatment in the case of PLS–DA and an N-fold cross-validation model was selected for this study. With this model, the individuals 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 validation. Thus, each object is used at least once in training and once in validation 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 3 were selected in all PLS–DA models. The search for marker phospholipids with discrimination capability was addressed with a web-based tool called ROCCET (ROC Curve Explorer & Tester, http://www.roccet.ca). This is a training and testing resource to build and validate biomarker models based on metabolomics data [13].

3. Results and discussion

3.1. Identification of phospholipids in serum samples

Phospholipids extracts obtained from human serum by SPE were analyzed by LC–QTOF in MS/MS acquisition mode. It is worth mentioning that the SPE cartridges used for sample preparation are packed with zirconia coated silica, which allows setting a selective interaction between the phosphate moiety (Lewis base) and Zr atoms, which act as a Lewis acid due to the acceptance of electrons by *d*-orbitals. The selectivity of this treatment enables to clean-up other lipids as well as proteins that could exert ionization suppression effects. Therefore, an oriented identification strategy was planned since only phospholipid metabolites were present in the extract.

Data-dependent methods based on product ion or neutral loss scanning modes in low-resolution mass spectrometry have been frequently tested for screening the different phospholipids classes and subclasses. For this purpose, characteristic ion products or neutral losses, which are representative of each phospholipid structure, have been monitored as identification and confirmatory strategy. However, mass accuracy has not been good enough to assure the identification of phospholipids, at least when using only one product ion or neutral loss scan modes or when working exclusively in one polarity mode. Despite of these evidences, few studies have described the utilization of high resolution mass spectrometry with QTOF or Orbitrap detectors to improve the identification capability of methods for analysis of phospholipids. However, the involvement of high-resolution mass spectrometry is crucial to success in metabolomics profiling of phospholipids as an attempt to find connection between their concentration in biofluids and certain pathologies.

Both positive and negative ionization modes were tested in this research to maximize the identification coverage. It is well-known that some phospholipid subclasses such as glycerophosphatidylcholines are optimally detected in positive ionization mode, while other subclasses such as glycerophosphatidylinositols and glycerophosphatidic acids are preferentially detected in the negative mode. This selection is supported on the sensitivity for ionization of precursor ions, but also on sensitivity and selectivity of product ions after suited fragmentation in the collision cell, which is crucial for identification. The preference of one or other ionization mode is strongly associated to the polar head group that defines the phospholipid subclass.

Fragmentation of phospholipids in mass spectrometry has been widely studied and the mechanisms involved in the fragmentation reactions of each phospholipid subclass have been described. Thus, representative product ions ascribed to the fragmentation of each phospholipid family are well-known [14]. This information was used in this research to develop an identification algorithm based on the most representative product ions pertaining to the different phospholipid families. For this purpose, human serum pools prepared with

aliquots of atherosclerotic patients affected by stable angina and NSTEMI were analyzed by the proposed approach combining SPE and LC–QTOF MS/MS. Glycerophosphatidylcholines and sphingomyelins, particularly ceramidephosphocholines, presented the same fragmentation pattern in positive ionization with the product ion *m/z* 184.0730 fitting the most characteristic fragment for both families, which corresponds to the phosphorylcholine moiety. In negative ionization mode, the fragmentation pattern for both families was also similar with the MS/MS spectra dominated by the presence of the $[M-CH_3]^-$ product ion. The main difference between the MS/MS spectra of these two phosphorylcholine classes was found in the product ions ascribed to the alkanoyl chains, acyl and sphingosine units, which aid decisively to the identification of them. Sphingomyelins with ceramidephosphoethanolamine structure were not detected in serum from atherosclerotic patients; therefore, this subclass was not monitored.

Glycerophosphatidylethanolamines were properly identified in both ionization modes by detection of representative neutral loss (*m/z* 141.0205 and 140.0096 in positive and negative ionization modes, respectively), which fits the phosphorylethanolamine moiety. In the case of glycerophosphatidylserines, the identification was supported on the detection of the neutral loss ascribed to the phosphorylserine polar group at 87.0350*m/z* in negative ionization mode, which fits the serine moiety. Fragmentation of phosphatidic acids was characterized by the product ion at 152.9975*m/z* in negative ionization mode, which is associated to the cyclic glycerophosphate derivative. The resting glycerophospholipid subclasses such as glycerophosphatidylinositols and glycerophosphatidylglycerols were not detected in serum samples of the selected cohort, despite their MS/MS behavior is also known.

Detection of the family-dependent product ions or neutral losses for each phospholipid subclass was combined with detection of the product ions associated to alkanoyl chains in negative ionization mode. This combination allowed elucidating the complete structure of the phospholipids by confirming the phospholipid subclass and the fatty acids present in the structure. Only in the

case of sphingomyelins the N-acyl chain did not provide a signal in the MS/MS spectra and, for this reason, identification of these phospholipids was based on the sphingophospholipid representative fragments and that corresponding to the unique alkanoyl chain.

The algorithm including the representative product ions, neutral losses of each phospholipid subclass, and the list of ions associated to the alkanoyl fragments for fatty acids from C12 to C30 was executed for all raw data sets generated after LC–QTOF MS/MS analysis of treated serum samples. Supplementary Table 1 shows the list of all phospholipids identified in human serum after analysis of the cohort studied in this research. Information about chromatographic retention time, precursor ion as well as the characteristic product ions supporting the identification are included. As can be seen, most of the phospholipids corresponded to phosphatidylcholines, which is the most common family of phospholipids detected in human serum. Thirty-four glycerophosphatidylcholines, including lyso forms, were identified in the analyzed cohort. Concerning fatty acids of phosphatidylcholine structure, saturated, monounsaturated and polyunsaturated fatty acids were detected. Additionally, fatty acids with odd number of carbons (C15:0 and C17:0) were detected, which could be explained by intake in the diet. It is also worth mentioning the detection of less conventional glycerophosphatidylcholines such as the plasmalogenlysoPC(P-16:0), lysoPC(OH-16:0) and PC(O-16:0/20:4), which contain palmityl alcohol in their structure.

Lysophosphatidylethanolamines with the most common fatty acids from 16:0 to 22:6 were also detected in human serum, including lysoPE(O-18:1), which contains oleyl alcohol. A profile formed by seven sphingomyelins as ceramidephosphocholines was also identified. Conventional fatty acids such as palmitic acid, stearic acid and oleic acid were present in the sphingomyelin structure, while the sphingosine unit consisted of saturated, monounsaturated and polyunsaturated amino alcohols (18:0, 16:1, 18:1, 18:2 and 24:2).

Two other phospholipid subclasses such as phosphatidylserines and phosphatidic classes were also identified in the cohort of atherosclerotic patients. The most concentrated LysoPA in serum, with oleic acid, was identified and detected in all samples. Finally, four phosphatidylserines including two plasmalogens such as PS(P-22:6/18:0) and PS(P-20:2/18:2), PS(O-20:0/22:2) with an arachidyl alcohol and PS(18:4/22:4) with two polyunsaturated fatty acids conjugated with glycerol.

3.2. Quality control and sample stability

A pool of serum samples from atherosclerotic patients was prepared for a quality control test and analysis of sample stability. Healthy control individuals were not included in this research since medication would be responsible of differences in the metabolite profiling. The extract obtained after application of the SPE protocol was sequentially analyzed 40 times by LC–TOF/MS in scan mode to obtain the profile of phospholipids present in serum. Ionization in positive mode was selected because glycerophosphocholines are the most concentrated phospholipid subclass in human serum, although the resting subclasses can also be detected in this polarity mode. The sample obtained after SPE treatment of the human serum pool was stored during the test in the autosampler of the LC system, which was refrigerated at 4 °C. Each analysis took 17 min to be completed, which means that the complete study was finished within 11.5 h. The initial data set was built after alignment of LC–TOF/MS chromatograms in scan mode according to retention time and mass accuracy. The tolerance window for both parameters was set at 0.4 min and 10 ppm for elution time and mass accuracy, respectively. Apart from this preprocessing step, a baselining treatment was also applied to the average of each variable across samples, which treats all molecular entities equally regardless of their abundance. After data pretreatment, the data set obtained was filtered using an algorithm based on detection frequency of entities in the replicates to minimize experimental variability. In particular, the algorithm eliminates the entities not detected in at least 100% of the replicates. A PCA was built for the data set obtained to evaluate the existence of variability sources that could be attributed

to instrumental precision in the detection step as well as to sample stability. The PCA scores plot is shown in Fig. 1 for PC1–PC2, which were only able to explain 58.7% of the total variability. As can be seen, the only variability source is detected after 24 replicates, while analyses 25–40 are clearly separated. Discrimination was observed along PC2, which only explained 25.8% of the total variability included in the data set. This variability source could be associated to sample alteration since the group of samples 25–40 are clearly separated. The loadings plot, shown in Supplementary Fig. 1, revealed that sphingomyelins were relevant to explain the differences observed between the two groups of samples. In fact, the concentration of sphingomyelins in the extract clearly decreased after 6 h. Taking into account that seven sphingomyelins were detected in the sample cohort, this aspect was considered in the processing of analysis sequences.

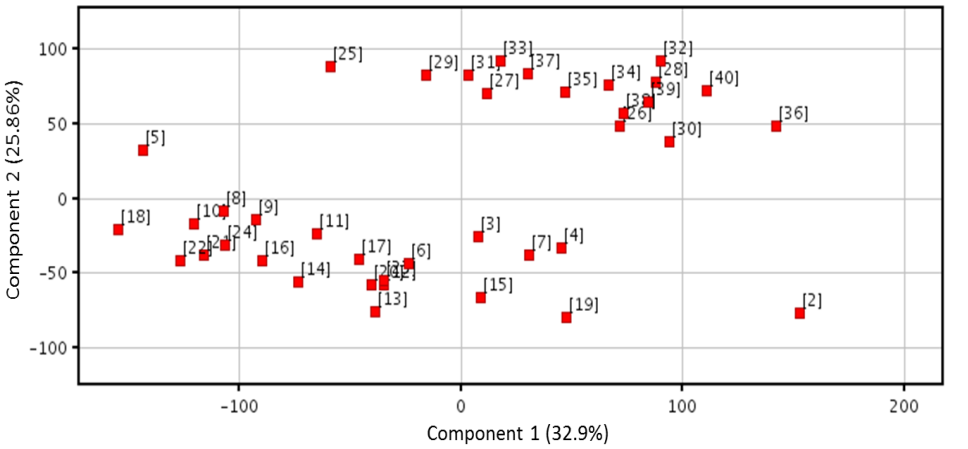


Figure 1. PCA scores plot for samples injected in the stability study.

The subsequent step was to estimate the standard deviation in the intensity of the phospholipids detected in the replicates 1–24. This estimation allowed concluding that at least 95% of the molecular entities presented in all replicates reported a coefficient of variation below 25%. Therefore, instrumental precision is quite consistent and the stability of the sample in terms of

phospholipids concentration can be ensured for 6 h at 4 °C (equivalent to 24 replicates).

3.3. Multivariate analysis for discrimination of atherosclerosis diagnostics

After identification, a database containing all compounds detected in the different serum samples was created. The database contained information about the molecular weight, formula, and retention time of each phospholipid. The algorithm "Find by database" was then executed to detect the presence of each phospholipid in the samples of the cohort and extract the quantitative area of each peak confirmed in the extracted ion chromatogram in MS mode. Confidence parameters were set at 10 ppm for mass accuracy and 0.3 min for retention time to minimize errors by quantitation of isomers. The resulting data set was then exported to the MPP software, which allowed filtering variables by frequency of occurrence in samples. Thus, the phospholipids no detected in at least 75% of samples in one condition (stable angina or NSTEMI) were excluded. The filtered data set was formed by 48 phospholipids distributed as follows: 13 phosphatidylcholines, 18 lysophosphatidylcholines, 6 sphingomyelines, 7 phosphatidylethanolamines, 3 phosphatidylserines and 1 phosphatidic acid, which cover practically the different subclasses of phospholipids. The filtering step based on the occurrence of phospholipids in serum samples allowed ensuring the representativity of the data set associated to the serum phospholipid profile of atherosclerotic patients. The phospholipids, which were present in the processed data set, are listed in Fig. 2.

Once completed the data set containing quantitative information in relative terms of phospholipids levels in human serum, the next step of this research was to find a potential connection of this data matrix with atherosclerosis diagnostic. Previously, the data set was pretreated by log transformation to minimize the relevance of the phospholipids concentration in serum. Supplementary Fig. 1 shows the effect of this transformation on the variability observed for each phospholipid.

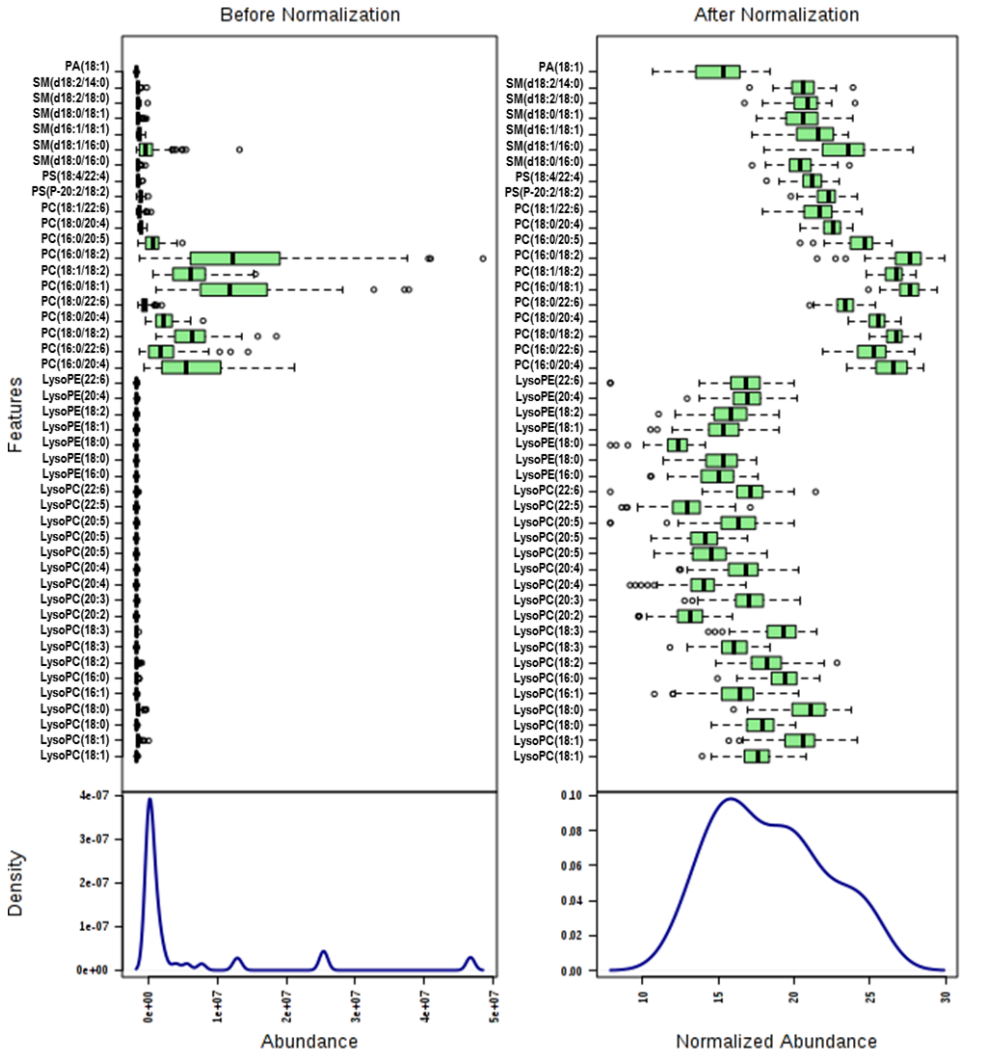


Figure 2. Phospholipids detected in all samples including their peak area variability before and after log2 transformation. The areas are plotted using box-and-whiskers.

Multivariate analysis was applied to develop a statistical model with capability to discriminate between the two groups diagnosed with atherosclerosis: stable angina *versus* NSTEMI. Thus, a PLS-DA model was created to find classification patterns related to the atherosclerosis diagnostic. The scores graph for the PLS–DA study is illustrated in Fig. 3. As can be seen,

discrimination was observed between the two groups using three-dimensional plots. Sensitivity and specificity values obtained in the training model were 76.1 and 76.2%, respectively. On the other hand, the stability of the model was assessed by cross-validation, which reported sensitivity and specificity values of 62.7 and 56.7%, respectively. These results suggest that phospholipid levels would be different in the two diagnostics of atherosclerosis under study, which means that the composition and concentration of phospholipids are altered by the occurrence or not of hypoxia [15-16]. For this reason, a prediction capability study is necessary to obtain further information.

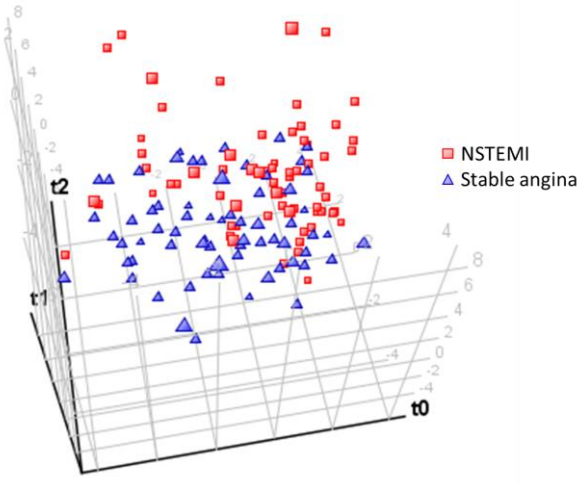


Figure 3. Scores plot of the PLS–DA model built to discriminate between patients diagnosed with stable angina and those diagnosed with non-ST elevation myocardial infarction (NSTEMI).

3.4. Prediction capability study, development and validation of a markers panel

Taking into account the discrimination level observed between the two atherosclerosis diagnostics in cardiovascular patients, it would be of interest to confirm the existence of phospholipid patterns responsible for the observed variability. For this purpose, the web-based tool ROCCET was used to find phospholipids with the highest contribution according to statistical analysis to

explain the occurrence of atherosclerotic diagnostics. In a first step of this protocol illustrated in Supplementary Fig. 2, identification of phospholipids statistically relevant between the two atherosclerosis diagnostics was carried out by the *t*-test with *p*-value<0.05 as significance level. Table 2 lists the four phospholipids that reported differences in serum concentration between stable angina and NSTEMI patients, which means discrimination attributed to the occurrence of hypoxia. The highest significance was observed for lysoPC(20:5), which contains the essential omega-3 eicosapentaenoic acid. Lysophosphatidylcholines have been previously related to atherosclerosis as they are the major phospholipid components of atherogenic lipoproteins, generated upon extensive LDL oxidation [17]. Two of the most concentrated phospholipids in human serum, PC(18:1/18:2) and PC(18:0/20:4), resulted significant (*p*-value 0.0459 and 0.0115, respectively) to explain the severity of the atherosclerosis diagnostic, stable angina or NSTEMI. The fatty acids present in these two glycerophospatidylcholines were the three most concentrated C18 fatty acids (stearic acid, oleic acid and linoleic acid), and arachidonic acid, an omega-6 essential fatty acid. The fourth phospholipid was the SM(d18:2/14:0), which gave a *p*-value of 0.0030.

Table 2. Phospholipids with high significante ($p<0.05$) for discrimination of patients with stable angina and those diagnosed with non-ST elevation myocardial infarction (NSTEMI).

Phospholipid	**AUC**	***t*-Test *p*-value**	**Fold change**
PC(20:5)	0.6842	0.0008	0.9198
SM(d18:1/16:0)	0.6142	0.0030	-0.8448
PC(18:0/20:4)	0.6330	0.0115	-0.2646
PC(18:1/18:2)	0.6090	0.0459	-0.2145

Supplementary Fig. 3 shows the box and whiskers diagrams obtained for the four metabolites in the two groups considered. As can be seen, only in the case of lysophosphatidylcholine levels are slightly increased in patients with NSTEMI in comparison with patients diagnosed with stable angina. This effect has been previously observed by Fernández et al. who concluded that lysoPC(16:0) and lysoPC(20:4) were positively correlated with HDL and, thus,

negatively correlated with cardiovascular disease risk factor [18]. Furthermore, polyunsaturated acyl lysophosphatidylcholine prevent inflammation [19]. The opposite behavior is shown by phosphatidylcholines and sphingomyelin, the levels of which are slightly higher in angina stable patients. This found is also in accordance with preceding research as plasma sphingomyelin levels have been found to be positively and independently related to the presence of coronary artery disease [20]. The same occurs with phophatidylcholines, which are related to inflammation and can be up synthetized in atherosclerotic lesions [21].

The four statistically significant phospholipids detected in stable angina and NSTEMI patients were selected as features subset to evaluate the predictive power between the two conditions under study. The discrimination model was based on PLS–DA and the cohort was randomly divided into training and validation sets, which included 80% and 20% of the samples of each condition (113 and 27 samples for each set), respectively. Prediction model parameters obtained with the PLS–DA model generated for the two classes in the training and validation sets are shown in Table 3. Thus, the disease prediction capability in the training set was 70.7 and 66.0% for stable angina and NSTEMI patients, respectively. On the other hand, the validation set by application to an independent cohort improved the predictive power for angina stable patients to 92.3%, while this was slightly decreased up to 50.0% for NSTEMI patients.

Table 3. Prediction capability values and false positive rates (expressed as percent) obtained for the validation and training model created with the four phospholipids selected for patients with stable angina versus those diagnoses with non-ST elevation myocardial infarction (NSTEMI).

	Stable angina prediction capabiliy	**NSTEMI prediction capability**	**False positive for NSTEMI**	**False positive for stable angina**
Training model	70.7	66.0	30.5	32.7
Validation model	92.3	50.0	36.9	12.5

This characterization was completed with the ROC analysis of the panel formed by the selected phospholipids. Fig. 4 illustrates the ROC curves obtained for this panel in the training and validation sets. As can be seen, the area under

the curve parameter (AUC) was 0.715 for the training set, while this was 0.747 for the external validation set. The performance of this panel of markers was assessed by estimation of sensitivity, selectivity and accuracy both in the training set and validation set by application of the model to this sub-cohort of samples not used to build the model. The optimum predictive model in the training set gave 66% and 70% of sensitivity and specificity, respectively. Nevertheless, the model could be improved in terms of sensitivity or specificity if a decrease of false positive or false negative rates is desired without reducing any parameter below 60%. Thus, the best conditions to minimize false positives provided 77% sensitivity and 60% specificity, while the best model targeted at specificity reported 76% for this parameter and 60% sensitivity. It is worth mentioning the stability of the model, since this is well-balanced for sensitivity and specificity. The application of the statistical model to the validation set allowed improving both parameters, which were 79 and 78% for sensitivity and specificity, respectively. This means that only two of ten patients would be erroneously diagnosed with NSTEMI by metabolomics analysis (false positives), which could be confirmed by use of an electrocardiogram.

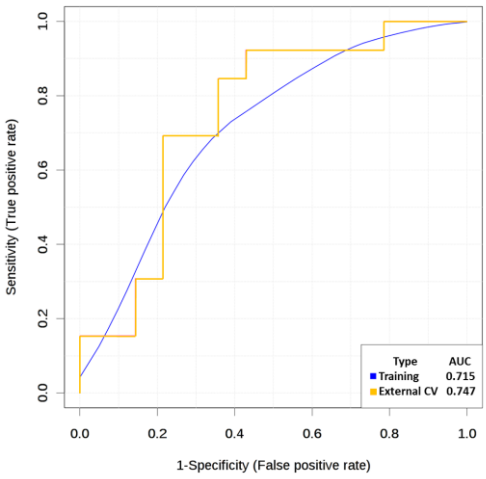


Figure 4. ROC curve obtained from the model created using the four phospholipids significantly different between patients with stable angina and those with NSTEMI. The ROC curve for both training and validation sets are represented by their area under the curve (AUC).

Finally, the predictive capability of the panel of markers improved the independent prediction power of the four target phospholipids. Fig. 5 shows the ROC curves obtained for each phospholipid with information about the optimum values of sensitivity and specificity. The best behavior was found for lysoPC(20:5) that provided 62% of sensitivity and 70% of specificity, while a noticeable decrease was observed for the other three phospholipids. These results justify the selection of the panel to aid in the diagnosis of stable angina and NSTEMI in atherosclerotic patients and provide information which could be used for catheterization. Concerning phospholipids with the highest significance to explain the variability observed, it is worth mentioning the presence of essential omega-3 and omega-6 fatty acids such as linoleic acid, arachidonic acid and eicosapentaenoic acid.

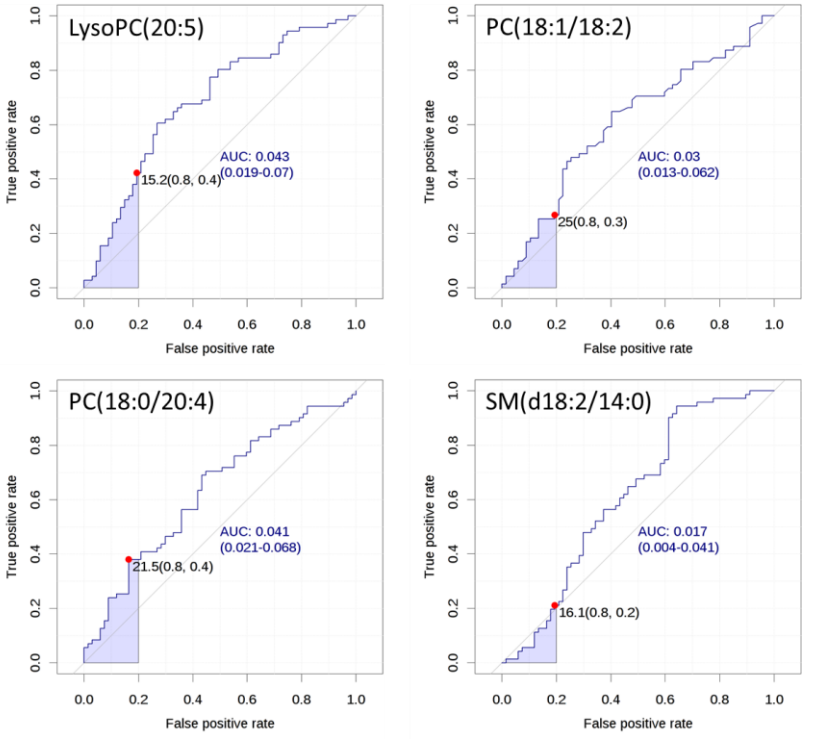


Figure 5. Individual ROC curve for lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0) when discriminated from stable angina patients and those diagnosed with NSTEMI.

4. Conclusions

The analysis of phospholipids serum profiles by LC–QTOF MS/MS has shown to be a suitable strategy for biomarker discovery. A group of four phospholipids significantly different between patients with stable angina and those diagnoses with NSTEMI have been used to build a panel of markers which were composed by lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0).

The panel of markers proposed for discrimination between stable angina and NSTEMI presented a disease prediction capability in the training set of 70.7 and 66.0% for stable angina and NSTEMI patients, respectively. Furthermore, the ROC curve obtained for the model gives an area under the curve parameter (AUC) was 0.715 for the training set, while it was 0.747 for the external validation set. This ROC curve shows a 66% and 70% of sensitivity and specificity, respectively. These results warrant a larger scale study to validate the proposed model and support the usefulness of LC–QTOF/MSMS strategy to profile phospholipids for development of tools to aid in clinical diagnostics.

Acknowledgements

The Spanish Ministerio de Ciencia e Innovación (MICINN) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2009-0734). F. Priego-Capote and M. Calderón Santiago are also grateful to the MICINN for a Ramón y Cajal Contract (RYC-2012-37428) and an FPU scholarship, respectively. The Miguel Servet Hospital (Zaragoza, Spain) is also thanked for donors serum samples.

References

[1] S. Petersen, V. Peto, M. Rayner, J. Leal, R. Luengo-Fernández, A. Gray, European Cardiovascular Disease Statistics, British Heart Foundation: London, U.K., 2005.

[2] C.M. Rimmerman, in: W.D. Carey (Ed.), Current Clinical Medicine, Saunders, Philadelphia 2010.

[3] E.M. Antman, M. Cohen, P.J.L.M. Bernink, C.H. McCabe, T. Horacek, G. Papuchis, B. Mautner, R. Corbalan, D. Radley, E. Braunwald, JAMA 284 (2000) 835.

[4] M.G. Barderas, C.M. Laborde, M. Posada, F. Cuesta, I. Zubiri, F. Vivanco, G. Alvarez-Llamas, J. Biomed. Biotechnol. 2011 (2011) 790132.

[5] K. Wang, R.W. Asinger, H.J. Marriott, N. Engl. J. Med. 349 (2003) 2128.

[6] R.W. Mahley, T.L. Innerarity, S.C. Rall, K.H. Weisgraber, J. Lipid. Res. 25 (1984) 1277.

[7] W.P. Castelli, R.J. Garrison, P.W. Wilson, R.D. Abbott, S. Kalousdian, W.B. Kannel, JAMA 256 (1986) 2835.

[8] A.Y. Tu, H.I. Nishida, T. Nishida, J. Biol. Chem. 268 (1993) 23098.

[9] G. Wolfbauer, J.J. Albers, J.F. Oram, Biochim. Biophys. Acta 143 (1999) 965.

[10] J.D. Bagdade, W.F. Buchanan, T. Pollare, H. Lithell, Atherosclerosis 117 (1995) 209.

[11] P. Xia, M.A. Vadas, P. Xia, M.A. Vadas, K.A. Rye, P.J. Barter, J.R. Gamble, J. Biol. Chem. 274 (1999) 33143.

[12] J.A. Ibdah, S. Lund-Katz, M.C. Philips, Biochemistry 28 (1989) 1126.

[13] J. Xia, D.I. Broadhurst, M. Wilson, D.S. Wishart, Metabolomics 9 (2013) 280.

[14] M. Pulfer, R.C. Murphy, Mass Spectrom. Rev. 22(5) (2003) 332.

[15] J. Jezková, O. Nováková, F. Kólar, E. Tvrzická, J. Neckár, F. Novák, Mol. Cell Biochem. 232(1-2) (2002) 49.

[16] J. McHowat, S. Liu, M.H. Creer, Am. J. Physiol. 274 (1998) C1727.

[17] C.A. Schaefer, C.R. Kuhlmann, C. Gast, S. Weiterer, F. Li, A.K. Most, T. Neumann, U. Backenköhler, H. Tillmanns, B. Waldecker, J. Wiecha, A. Erdogan, Vascul. Pharmacol. 41(2) (2004) 67.

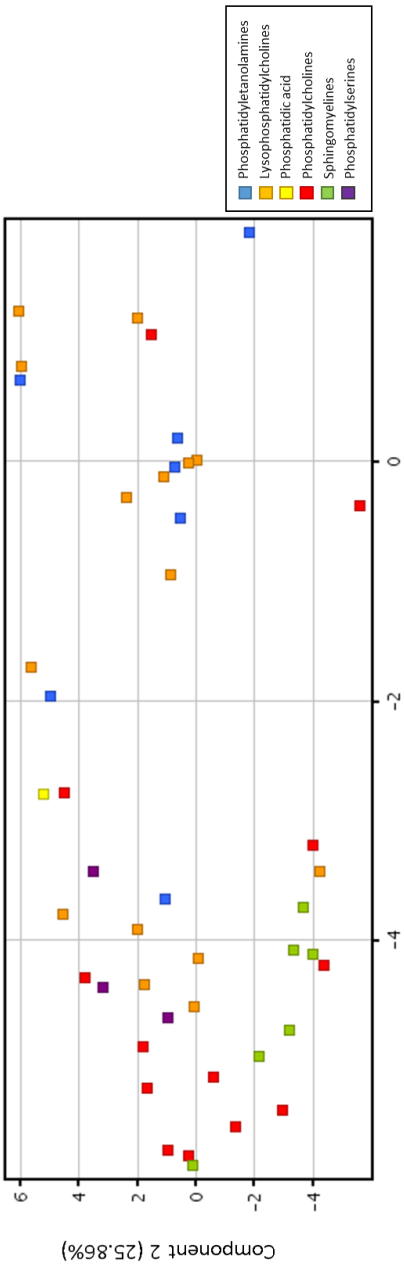
[18] C. Fernández, M. Sandin, J.L. Sampaio, P. Almgren, K. Markiewicz, M. Hoffmann, T. Hedner, B. Wahlstrand, K. Simons, A. Schevcherko, P. James, O. Melander, PLOS one 8 (2013) e71846.

[19] N.D. Hung, D.E. Sok, M.R. Kim, Inflamm. Res. 61 (2011) 473.

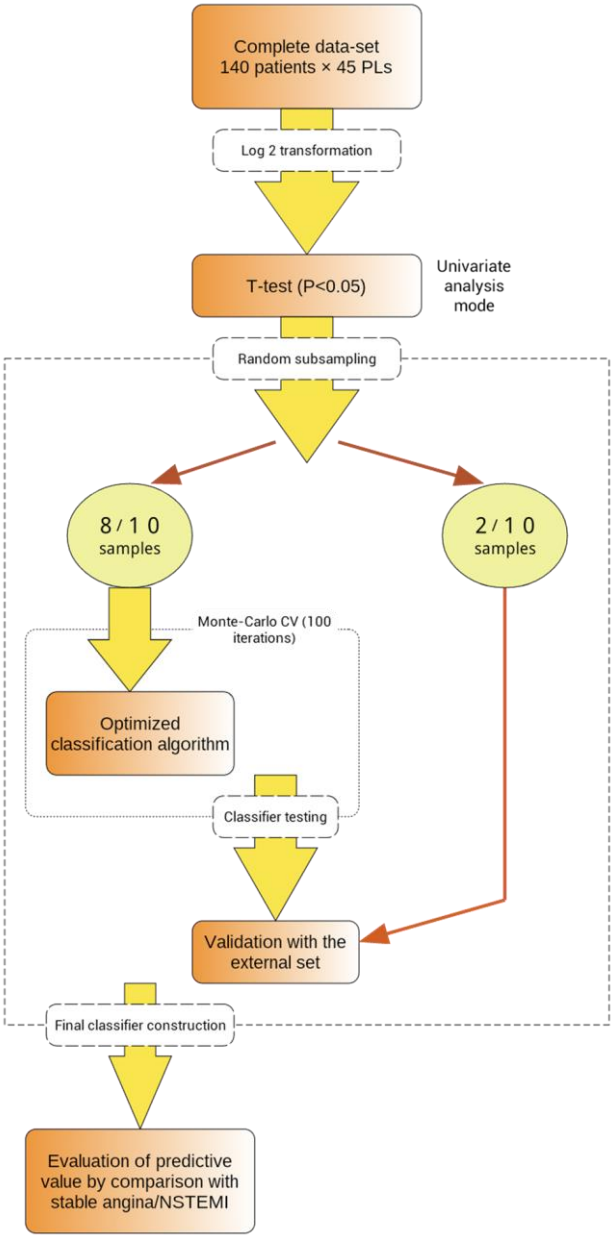
[20] X.C. Jiang, F. Paultre, T.A. Pearson, R.G. Reed, C.K. Francis, M. Lin, L. Berglund, A.R. Tall, Atherioscler. Thromb. Vasc. Biol. 20(12) (2000) 2614.

[21] Y. Shiratori, M. Houweling, X. Zha, I. Tabas, J. Biol. Chem. 270 (1999) 29894.

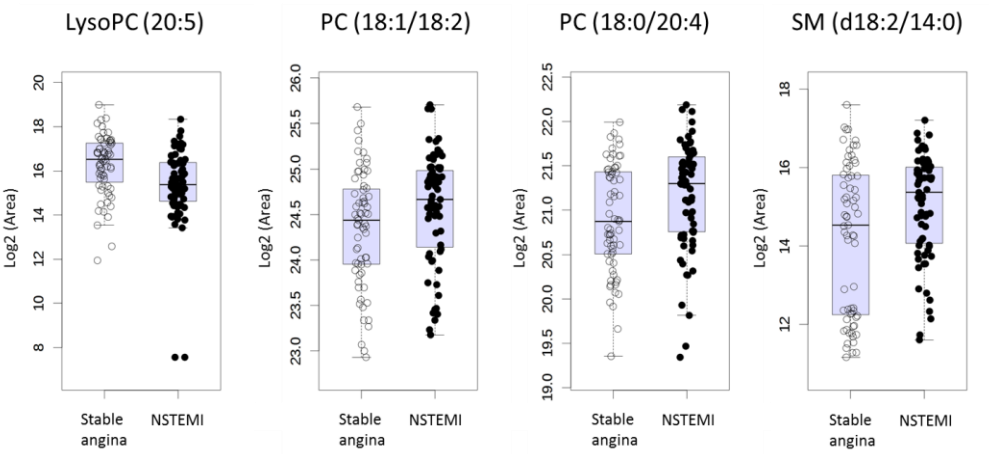
Supplementary material



Supplementary Figure 1. PCA loadings plot for samples injected in the stability study.



Supplementary Figure 2. Schematic overview of the procedure for generation and validation of the classification model.



Supplementary Figure 3. Box-and-whisker obtained from the log2 of the areas of lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0) in the two groups under study: patients diagnosed with stable angina and those diagnosed with NSTEMI.

Supplementary table 1. Phospholipids identified in human serum by the method based on HybridSPE® cartridges and LC–QTOF MS/MS.

Compound name	Formula	Mw	RT (*min*)	Precursor ion (*m/z*)		Familiar characteristic product ion (*m/z*)		Family characteristic neutral loss (*m/z*)		Fatty acid chain characteristic product ion (*m/z*)
				Positive ionization mode	Negative ionization mode	Positive ionization mode	Negative ionization mode	Positive ionization mode	Negative ionization mode	Negative ionization mode
Phosphatidylethanolamines (PE)										
LysoPE(16:0)	C21H44NO7P	453.2868	12.1 12.3	454.2942	452.2786	313.2744	-	141.0198	-	255.2335
LysoPE(16:1)	C21H42NO7P	451.2724	11.2	452.2795	450.2629	311.2554	-	141.0241	-	253.2152
LysoPE(18:0)	C23H48NO7P	481.3177	14.6 14.2	482.3250	480.3103	341.3065	196.0401	141.0185	-	283.2625
LysoPE(18:1)	C23H46NO7P	479.3012	12.7 12.9	480.3082	478.2940	339.2885	-	141.0197	-	281.2483
LysoPE(18:2)	C23H44NO7P	477.288	11.6 11.8	478.2948	476.2786	337.2738	-	141.0210	-	279.2321
LysoPE(20:3)	C25H46NO7P	503.3001	12.4	504.3068	502.2950	363.2893	196.0377	141.0175	-	305.2484
LysoPE(20:4)	C25H44NO7P	501.2837	11.6 11.8	502.2939	500.2805	361.2747	196.0339	141.0192	-	303.2338 259.2435
LysoPE(22:6)	C27H44NO7P	525.2856	11.7 11.8	526.2922	524.2789	385.2718	196.0337	141.0204	-	327.2339 283.2413
LysoPE(O-18:1)	C23H48NO6P	465.3219	15.4	466.3301	464.3158	325.3114	196.0358	141.0187	-	-

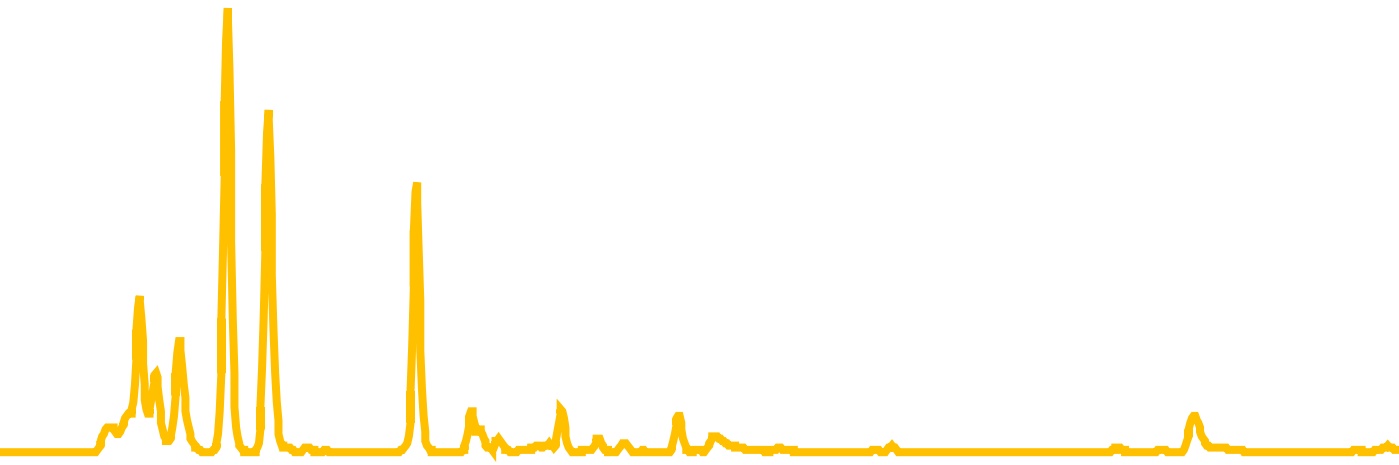
Phosphatidylcholines (PC)										
LysoPC(P-16:0)	$C_{24}H_{50}NO_6P$	479.3376	13.2	480.3438	524.3390	184.0733 104.1077	464.3154	-	60.0236	239.2324
LysoPC(14:0)	$C_{22}H_{46}NO_7P$	467.3012	10.4 10.7	468.3082	512.3001	184.0728 104.1069	452.2757	-	60.0244	227.2007
LysoPC(15:0)	$C_{23}H_{48}NO_7P$	481.3168	11.3 11.6	482.3242	526.3141	184.0731 104.1073	466.2922	-	60.0219	241.2171
LysoPC(16:0)	$C_{24}H_{50}NO_7P$	495.3325	12.2 12.5	496.3404	540.3321	184.0732 104.1070	480.3104 224.0694	-	60.0217	255.2335
LysoPC(16:1)	$C_{24}H_{48}NO_7P$	493.3168	10.9 11.2	494.3246	538.3152	184.0728 104.1072	478.2930	-	60.0222	253.2155
LysoPC(17:0)	$C_{25}H_{52}NO_7P$	509.3481	12.5	510.3553	554.3471	184.0727 104.1068	494.3244	-	60.0227	269.2470
LysoPC(17:1)	$C_{25}H_{50}NO_7P$	507.3325	12.1	508.3761	552.3302	184.1073	492.3096	-	60.0206	267.2313
LysoPC(18:0)	$C_{26}H_{54}NO_7P$	523.3638	14.3 14.6	524.3732	568.3627	184.0734 104.1071	508.3412 224.0701	-	60.0215	283.2636
LysoPC(18:1)	$C_{26}H_{52}NO_7P$	521.3481	12.7 12.9	522.3564	566.3469	184.0727 104.1064	506.3249 224.0689	-	60.0220	281.2497
LysoPC(18:2)	$C_{26}H_{50}NO_7P$	519.3325	11.6 11.9	520.3405	564.3304	184.0732 104.1073	504.3090 224.0695	-	60.0214	279.2337
LysoPC(18:3)	$C_{26}H_{48}NO_7P$	517.3168	10.9 11.1	518.3205	562.3185	184.0722 104.1076	502.2931	-	60.0254	277.2161
LysoPC(20:2)	$C_{28}H_{54}NO_7P$	547.3638	13.6	548.3745	592.3575	184.0734 104.1080	532.3412	-	60.0163	307.2631
LysoPC(20:3)	$C_{28}H_{52}NO_7P$	545.3481	12.3 12.5 14.5 14.7	546.3538	590.3473	184.0730 104.1055	530.3246	-	60.0227	305.2481

LysoPC(20:4)	$C_{28}H_{50}NO_7P$	543.3325	11.7 11.9 12.8 13.1	544.3406	588.3317	184.0729 104.1063	528.3107	-	60.0210	303.2320
LysoPC(20:5)	$C_{28}H_{48}NO_7P$	541.3168	10.9 11.1 11.6 11.9	542.3217	586.3154	184.0740 104.1076	526.2940	-	60.0214	301.2171
LysoPC(22:4)	$C_{30}H_{54}NO_7P$	571.3638	13.3 13.1	572.4512	616.3620	184.0735	556.3402	-	60.0218	331.2605
LysoPC(22:5)	$C_{30}H_{52}NO_7P$	569.3481	12.3	570.3573	614.3462	184.0727 104.1079	554.3242	-	60.0220	329.2515
LysoPC(22:6)	$C_{30}H_{50}NO_7P$	567.3325	11.9 11.8	568.3403	612.3311	184.0731 104.1072	552.3101	-	60.0210	327.2332
LysoPC(hydroxy-16:0)	$C_{24}H_{50}NO_8P$	511.3274	8.4 10.9	512.3341	556.3269	184.0741 104.1081	496.3038	-	60.0231	271.2269
PC(16:0/18:3)	$C_{42}H_{82}NO_8P$	755.5465	15.3	756.5539	800.5424	184.0738	740.5210	-	60.0214	255.2324 277.2177
PC(16:0/18:2)	$C_{42}H_{80}NO_8P$	757.5622	16.5	758.5693	802.5611	184.0731	742.5395 224.0715	-	60.0216	255.2336 279.2337
PC(O-16:0/20:4)	$C_{44}H_{82}NO_8P$	767.5828	15.5	768.5878	812.5801	184.0733	752.5575	-	60.0226	303.2344
PC(16:0/20:5)	$C_{44}H_{78}NO_8P$	779.5465	16.2	780.5537	824.5473	184.0735 104.1077	764.5253 224.0678	-	60.0220	255.2335 301.2163
PC(16:0/20:4)	$C_{44}H_{80}NO_8P$	781.5621	15.8	782.5684	826.5599	184.0734	766.5404 224.0712	-	60.0195	255.2320 303.2337
PC(16:0/18:1)	$C_{42}H_{82}NO_8P$	759.5778	16.1	760.5847	804.5751	184.0727	744.5531	-	60.0220	255.2328 281.2479
PC(18:1/18:2)	$C_{44}H_{82}NO_8P$	783.5778	15.8	784.5836	828.5725	184.0733	768.5524 224.0716	-	60.0201	279.2316 281.2527

PC(18:0/18:2)	C44H84NO8P	785.5934	16.4	786.6007	830.5918	184.0734	770.5712 224.0724	-	60.0206	279.2343 283.2643
PC(16:0/22:6)	C46H80NO8P	805.5621	15.7	806.5690	850.5621	184.0730	790.5421 224.0685	-	60.0200	255.2317 327.2283
PC(16:0/22:5)	C46H82NO8P	807.5778	15.7	808.5825	852.5754	184.0734	792.5554	-	60.0200	255.2278 329.2481
PC(18:1/20:4)	C46H82NO8P	807.5778	15.1	808.5829	852.5752	184.0732	792.5556	-	60.0196	281.2419 303.2309
PC(18:0/20:4)	C46H84NO8P	809.5935	16.3	810.6006	854.5902	184.0731	794.5720 224.0695	-	60.0182	283.2628 303.2350
PC(18:0/20:3)	C46H86NO8P	811.6091	16.5	812.6157	856.6086	184.0734	796.5823	-	60.0263	283.2630 305.2489
PC(18:1/22:6)	C48H82NO8P	831.5778	15.7	832.5823	876.5883	184.0742	816.5500	-	60.0383	281.2499 327.2311
PC(18:0/22:6)	C48H84NO8P	833.5935	15.7	834.5991	878.5939	184.0738	818.5707 224.0710	-	60.0232	283.2614 327.2312
Sphingomyelines (SM)										
SM(d18:0/16:0)	C39H81N2O6P	704.5832	15.1	705.5828	749.5703	184.0739	689.5528	-	60.0175	255.2320
SM(d18:1/16:0)	C39H79N2O6P	702.5676	15.9	703.5745	747.5657	184.0732	687.5431	-	60.0226	255.2313
SM(d16:1/18:1)	C39H77N2O6P	700.519	14.8	701.5585	745.5476	184.0721	685.5301	-	60.0175	281.2502
SM(d18:0/18:1)	C41H83N2O6P	730.5989	15.6	731.6052	775.5907	184.0715	715.5736	-	60.0171	281.2446
SM(d24:2/18:0)	C45H85N2O6P	812.6771	14.8	813.6032	857.6786	184.0739	797.6507	-	60.0279	283.2573
SM(d18:2/18:0)	C41H81N2O6P	728.5823	15.8	729.5897	773.5808	184.0732	713.5532	-	60.0276	283.2625
SM(d18:2/14:0)	C37H73N2O6P	672.5206	15.2	673.5279	717.5185	184.0722	657.4978	-	60.0207	227.2011
Phosphatidylserines (PS)										
PS(P-22:6/18:0)	C46H80NO9P	821.5571	15.5	822.5652	820.5593	-	733.5187	-	87.0406	283.2513
PS(P-20:2/18:2)	C44H78NO9P	795.5414	16.0	796.5506	794.5406	-	707.5113	-	87.0293	279.2339
PS(O-20:0/22:2)	C48H92NO9P	857.6510	15.7	858.6575	856.6319	-	769.5985	-	87.0334	335.2987
PS(18:4/22:4)	C46H74NO10P	831.5050	16.1	832.5123	830.4975	-	743.4669	-	87.0306	275.2015 331.2636
Phosphatidic acid (PA)										
LysoPA(18:1)	C21H41O7P	436.2590	8.5	437.2660	435.2526	155.0099	152.9975	-	-	281.2512

Chapter 9:

Determination of essential amino acids in human serum by a targeting method based on automated SPE–LC–MS/MS: Discrimination between atherosclerotic patients



Journal of Pharmaceutical and Biomedical Analysis

70 (2012) 476–484

Determination of essential amino acids in human serum by a targeting method based on automated SPE–LC–MS/MS: Discrimination between atherosclerotic patients

M. Calderón-Santiago [a,b], *F. Priego-Capote* [a,b], *J.G. Galache-Osuna*[c], *M. D. Luque de Castro*[a,b]*

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Department of Cardiology, Hemodynamic and Interventional Cardiology, University Hospital Miguel Servet, Zaragoza, Spain.*

Determination of essential amino acids in human serum by a targeting method based on automated SPE–LC–MS/MS: Discrimination between atherosclerotic patients

*M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M. D. Luque de Castro**

Abstract

An automated method based on a hyphenated SPE–LC–MS/MS configuration has been optimized for the determination of essential amino acids (threonine, valine, methionine, leucine, lysine, tryptophan, and phenylalanine) in human serum, with the aim of discriminating between different states of coronary artery disease. Validation in terms of sensitivity (detection limits below 28.0 ng on column) and precision (repeatability expressed as relative standard deviation below 6.0%) supports the suitability of the method for application to a cohort of 122 atherosclerosis patients confirmed catheterization test. The cohort was composed by 80 individuals diagnosed with stable angina and 42 patients who suffered from acute myocardial infarction (AMI). Both groups of individuals are differentiated by the occurrence of ischemia in AMI patients due to the formation of thrombi. The chemometric treatment of the data obtained by multivariate analysis of variance (MANOVA) allowed by isoleucine, a comparison both groups of diagnosed patients. Therefore, amino acids whose serum levels were affected by ischemia have been identified. The contribution of risk factors such as obesity and hypercholesterolemia as well as the individuals' sex to the concentration of essential amino acids has also been studied.

1. Introduction

Amino acids are molecules with a common structure that contains an amine group, a carboxylic acid group and a characteristic side-chain that varies among amino acids. They have critical functions to life with a direct involvement in metabolism by serving as the building blocks of proteins. Apart from their functions in proteins synthesis, amino acids are key regulators of gene expression and the protein phosphorylation cascade, and act as precursors for synthesis of hormones and low-molecular weight nitrogenous substances with enormous biological importance such as nitric oxide, polyamines, glutathione, taurine, etc [1]. The biochemical relevance of amino acids, their levels and those of their metabolites may be of interest in the study of metabolic and nutritional disorders as well as in other pathologies related with oxidative stress.

The heart is one of the organs more affected by reactive oxygen species (ROS) [2], so one of the pathologies related with oxidative stress is atherosclerosis, a disease that represents a state of heightened oxidative stress characterized by lipid and protein oxidation in the vascular walls [3–5]. Furthermore, atherosclerosis is clearly related with the inflammatory cascade where ROS are key mediators starting from the initiation of fatty streak development, through lesion progression, to ultimate in plaque rupture and thrombi formation (resulting in myocardial infarction) [6].

Atherosclerosis causes are not clear, but there are certain traits, conditions or habits known as risk factors that may raise the risk of suffering atherosclerosis [7]. Most common atherosclerosis risk factors are age, hypercholesterolemia, hypertension, diabetes, obesity and smoking. Among the different factors, obesity is a major public health crisis worldwide considered a leading risk factor for different pathologies such as diabetes, atherosclerosis, stroke, hypertension and some types of cancer [1,8,9]. Sex also influences atherosclerosis development, as men are more likely than women to develop this disease; however, sex difference narrows as men and women grow older.

Due to their central role in biochemistry, amino acids are important in nutrition. Plasma amino acid concentrations represent the balance of protein turnover (protein synthesis and degradation), amino acids absorption from diet, and metabolism of individual amino acids [9]. In fact, there are essential amino acids (*viz.* leucine, isoleucine, lysine, methionine, phenylalanine, threonine, histidine, tryptophan and valine) that the human body cannot synthesize at the concentration demanded for normal development and may be obtained from diet. Some essential amino acids have been proposed as cardiovascular disease biomarkers. Thus, threonine and phenylalanine have been proposed as myocardial infarction biomarkers [10]. The explanations given to this behavior are either their metabolism pathways are obstructed by the ischemia and oxygen deficiency condition, or the catabolism of the injured myocardial proteins to amino acids is enhanced due to the ligation of the left anterior descending coronary artery. Also, leucine, isoleucine and threonine are compounds affected immediately (10 min) after the onset of myocardial infarction [11].

Different analytical platforms have been proposed to determine amino acids in biofluids such as serum [12,13] and urine [14]. Most of them involve an electrophoretic or chromatographic step with subsequent determination by fluorimetry (usually with laser as excitation source), UV absorption spectrophotometry or mass spectrometry. A derivatization step prior to separation to enhance detection of amino acids by increasing signal given by analytes or ionization properties can also be involved. In fact, different labelling reagents such as 9-fluorenylmethyl chloroformate have been proposed for derivatization of amino acids [15]. Methods without previous derivatization have also been reported for determination of amino acids in serum and urine [16,17]. Furthermore, a Hitachi analyzer based on ion exchange chromatography with a spectrophotometer detector specific for amino acids is available in the market [18]. This analyzer, which allows the determination of a wide range of amino acids in 70 min, has been used to study the levels of amino acids in both patients with heart failure and normal individuals.

Because of the lack of automated methods for the simultaneous determination of essential amino acids of interest in the clinical field, the aims of this research were as follows: (i) to develop a method for determination of these amino acids in serum by on-line coupling of an automated solid-phase extraction system (Symbiosis-Pharma system) with an LC–MS/MS device; (ii) to apply the resulting method to the determination of the target compounds in a cohort formed by individuals affected by atherosclerosis; (iii) to correlate levels of essential amino acids in atherosclerotic patients and different risk factors such as obesity, sex, hypercholesterolemia and smoking habit.

2. Material and methods

2.1. Reagents

A multistandard solution of amino acids from Sigma–Aldrich (Madrid, Spain) with an individual concentration for each analyte of 0.5 mM ± 4% in a 0.2 N litium citrate buffer (pH 2.2), 2% of thiodiglycol and 0.1% of phenol was used to optimize both chromatographic separation and detection. L-valine, L-leucine, L-phenylalanine and L-lysine standards and phosphate buffer solution (PBS) were provided by Sigma–Aldrich. D-tryptophan was provided by Fluka (Spain) and L-methionine and L-threonin were from Merck (Madrid, Spain). An aqueous stock solution containing 100 μg/mL of each target amino acid was used to prepare diluted working solutions.

LC–MS grade acetonitrile and ammonium formate were purchased from Scharlab (Barcelona, Spain). Deionized water (18 mΩ cm) from a Millipore Milli-Q water purification system was used for preparation of all aqueous solutions.

2.2. Blood extraction and serum isolation

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated for 30 min at room temperature to allow coagulation. Then, the tubes

were centrifuged at 2000 g for 15 min at 4 °C to isolate the serum fraction (processing within 2 h after collection). Serum was placed in plastic ware tubes and stored at −80 °C until analysis.

2.3. Cohort selected for the study

A total of 122 patients clinically diagnosed with an episode of either stable angina (80 individuals) or myocardial infarction (42 individuals) formed the cohort. All patients were affected by atherosclerosis after evaluation through a cardiac catheterization. The main characteristics of the patients are shown in Table 1. The cohort was composed by individuals with an average age of 66±14, 81% of them male individuals, 53% smokers, 26% diabetic, 41% obese, 58% with hypertension and 50% with a high cholesterol level.

Table 1. Features of the cohort under study.

Characteristic	Patients (n=122)
Age	65.5 ± 13.5
Male sex, n (%)	99 (81.1)
Obesity, n (%)	50 (41.0)
High cholesterol level, n (%)	61 (50.0)

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, which were supervised by specialized personnel from Miguel Servet Hospital (Zaragoza, Spain). Individuals selected for this study were previously informed to obtain consent.

2.4. Instruments

Hyphenated SPE was performed with an automated robotized workstation Symbiosis-Pharma (Spark Holland, Emmen, The Netherlands) equipped with an autosampler Reliance (Spark Holland) furnished with a 250-μL sample loop and a refrigerated stacker sample compartment. The SPE workstation is endowed with a unit for SPE cartridge exchange —automatic cartridge exchanger (ACE)— and two high-pressure dispensers (HPD) for SPE

solvents delivery. Peek tubes of 0.25 mm i.d. (VICI, Houston, TX, USA) were used for all connections. The ACE unit included three switching valves, one of them equipped with a T-rotor, and two clamps. The Sparklink 3.10 SP#3 software was used to control the system. Hysphere MM anion cartridges (8 μm, 10 mm x 2.0 mm, Spark Holland) were used as sorbent material in the SPE step.

Chromatography was performed with an Agilent (Palo Alto, CA, USA) 1200 Series chromatograh composed by a binary pump and a Luna (3 μm, 100 mm x 4.6 mm) hydrophilic interaction chromatography column (HILIC) from Phenomenex (Torrance, CA, USA).

Detection was carried out by an Agilent 6410 Triple quadrupole mass spectrometer, furnished with an electrospray ionization (ESI) source.

2.5. Sample treatment

Serum samples (100 μL) immersed in an ice bath were treated for deproteinization with 600 μL methanol, the most common solvent used with this aim. The mixture was shaken for 1 min and the precipitate removed after centrifugation for 5 min at 6 °C and 13,800 × g. 400 μL of the upper phase was collected in a vial and diluted with 1 mL methanol, then placed in the Symbiosis autosampler, programmed to maintain the analytical sample at 15 °C until injection into the automatic SPE system coupled with the LC–MS/MS equipment.

2.6. Analysis protocol for amino acids determination in serum

Basically, the sample preparation step starts by activation of the sorbent phase with acetonitrile and subsequent equilibration before sample loading with 1 mL 50% acetonitrile containing 1% formic acid. After filling the 250-μL loop with analytical sample, this is loaded into the SPE cartridge with 0.5 mL 90% acetonitrile containing 1% formic acid. Under these conditions, the amino acids were retained in the cartridge, which was rinsed with 1 mL 70% acetonitrile to remove potential interferents. All these steps are performed in the left clamp.

Then, the cartridge is moved to the right clamp and 450 μL 90% acetonitrile containing 5% ammonium hydroxide passes through the cartridge for 2 min and 15 s to elute the amino acids. The eluate from the cartridge is merged with the chromatographic mobile phase by a T-rotor and both mixed phases go to the column. The elution time was 2 min 16 s, after which, the valve with the T-rotor is switched to allow the passage of mobile phase exclusively through the column. This initial mobile phase, with a flow of 0.6 mL/min, is composed by 90% acetonitrile with 5 mM of ammonium formate as ionization agent. The SPE step finishes by rinsing the cartridge consecutively with 1% formic acid, acetonitrile and water (1 mL each). The cartridges can be reused for three times without loss of efficiency.

The gradient used for LC separation of the compounds goes from 90% acetonitrile (maintained for 5 min) to 70% of acetonitrile in 10 min (also maintained for 5 min). A post time of 5 min was necessary to equilibrate the column prior to the next injection. The column temperature was set at 15 °C.

MS detection was performed in positive ESI mode at unit resolution in both quadrupoles. The ESI parameters were set as follows: 4 kV capillary voltage, 350 °C source temperature, and 50 psi pressure nebulizer. Nitrogen gas flowed at 9 mL/min to dry the eluent at the ionization source. Precursor and product ions selected for selected reaction monitoring (SRM), and those respective collision energies, are shown in Table 2.

The dwell time was set at 50 ms for all transitions. The entire analytical process was completed in 27 min. Nevertheless, the automated system enabled the chromatographic analysis of one sample while a second sample can be simultaneously prepared. The voltage of the first quadrupole was set at 90 V, which allows an efficient filtering for all selected precursor ions.

2.7. SRM-based quantitation

Calibration curves were run by diluting the stock solution with PBS. Calibration levels were expressed as ng on colum, ranging from 0.05 to 750. Five of them were injected in duplicate. This calibration model was selected because

no matrix effects were observed when recovery and ionization suppression were studied.

Table 2. Parameters used for LC–MS/MS determination of essential amino acids.

Compound	Retention time (min)	Precursor ion → Product ion *(m/z)* → *(m/z)*	Collision energy (V)
Phenylalanine	7.8	166.2 → 120.1	5
Leucine/isoleucine	9.0	132.2 → 86.1	5
Methionine	9.2	150.2 → 132.8	0
Tryptophan	9.5	205.1 → 188.1	5
Valine	10.7	118.2 → 72.3	10
Threonine	13.9	120.1 → 74.2	0
Lysine	14.3	147.2 → 84.1	10

3. Results and discussion

3.1. Development of the LC–MS method

A commercial standard solution containing the target amino acids (phenylalanine Phe, leucine Leu, isoleucine Ile, methionine Met, tryptophan Trp, valine Val, threonine Thr and lysine Lys) was used to optimize the LC–MS method. The process was initiated by optimization of the MS parameters, for which a generic chromatographic protocol starting with 90% acetonitrile and finishing with 40% acetonitrile was used. The recommendation of using a buffer solution when working with a HILIC column to obtain reproducible resolution results led to utilize 5 mM ammonium formate as ionizing agent, which gives a pH of 6. This value is in between the two pKa values of the majority of amino acids, at which the net charge of them is neutral.

Both positive and negative ionization modes were tested, but the highest sensitivity for most compounds was clearly achieved with the positive ESI mode. The electrospray variables were optimized by a univariate design. The ranges studied for temperature, pressure, and capillary voltage were 200–350 °C, 30–50 psi, and 3000–4000 V, respectively. The effect of these three variables on

the signal intensity corresponding to the target ions was positive, but not significant; so all the variables were established at their maximum values tested. Superior values led to an increase of background noise. The most abundant ions of all amino acids under the optimum conditions were established as precursor ions, as shows Table 2, which corresponded in all cases with the ion $[M+H]^+$.

The fragmentation of the different amino acids was studied using collision energies from 0 to 45 eV. Table 2 shows all product ions selected for each amino acid with the respective optimum collision energies. In most cases the product ion is produced after the loose of the carboxylic group, generating the ion $[M–COOH]^+$.

Different gradients were tested seeking for appropriate separation between the target analytes in a time as short as possible. All the assayed gradients involved acetonitrile and water both contaning 5 mM ammonium formate as ionizing agent. The pH of the mobile phase strongly affected the chromatographic separation; thus, pHs between 3 and 9 were tested, being pH 6 the optimum to attain a suited interaction with the chromatographic column. Values of pH out of the 5–7 range led to irreproducible results due to an inefficient retention of the analytes. Other variables that influenced the chromatographic separation —temperature and flow rate— were also studied. The best separation was obtained with the program under experimental.

3.2. Development of the SPE protocol

Optimization of the SPE step was carried out under real conditions by using a serum pool spiked with the target amino acids. 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 the process (detailed below under Section 3.3).

3.2.1. Selection of the suited sorbent material

Attending to the variability in the chemical structure of the essential amino acids, seven types of SPE sorbent with different retention properties were

tested: Hysphere CN (silica-based cyanopropyl phase), Hysphere C2 (silica-based ethyl phase), Hysphere C8 (silica-based octyl phase), Hysphere C18 (silica-based octadecyl phase), Resin GP (polymeric polydivinylbenzene phase), MM anionic (mixed-mode phase containing a strong anion exchange functional group), and MM cationic (mixed-mode phase containing a strong cationic exchange functional group). Since the phase behavior is different for each type of cartridge, three protocols were developed to test sorbents under conditions dependant on their retention mechanisms: one for the anionic sorbent, other for the cationic sorbent and a third one, a generic protocol, for the rest of the phases. The differences were the solution used for equilibration of the cartridge and sample loading, and that one for elution of the target amino acids. Thus, the pH played a crucial role in this preliminary experiment. The three protocols were initiated with a step for solvation of the sorbents by using 1 ml acetonitrile. The sample was loaded in the ionic protocols with a 20% acetonitrile (v/v) solution with a difference in the pH to enhance the ionic interactions. Thus, the pH of the loading solution was alkaline for the anionic sorbent by using 2% ammonia, and acid for the cationic sorbent by using 1% formic acid. On the other hand, the generic protocol was based on sample loading with water. The elution, programmed with the chromatographic mobile phase (90%, v/v, acetonitrile), was also distinctive of the three protocols in terms of pH. Thus, the elution from the anionic sorbent was performed with 5 mM ammonium formate at pH 4, while the same composition but at pH 6 was applied to the elution in the generic protocol. These protocols were compatible with the chromatographic process since the elution solution does not alter the separation. In fact, in the generic protocol the elution and mobile phases were identical, while minimal changes were used in the elution solution tested in the anionic protocol. On the other hand, the alkaline elution from the cationic sorbent was carried out by using 5% ammonia in the mobile phase. This pH change was not compatible with the chromatographic process. For this reason, an especial mode designed for elution with the minimum volume at a reduced flow-rate was employed to minimize the influence on the chromatographic separation. This mode, called focusing elution,

has been previously employed in clinical analysis for determination of thyronamines in human plasma and tissue [19] and cocaine and its metabolites in blood or urine [20,21]. This strategy would allow elution under alkaline conditions to be compatible with a chromatographic process in acid media. The instrumental configuration for this approach is shown in Fig. 1, which illustrates the main steps involved in the SPE process.

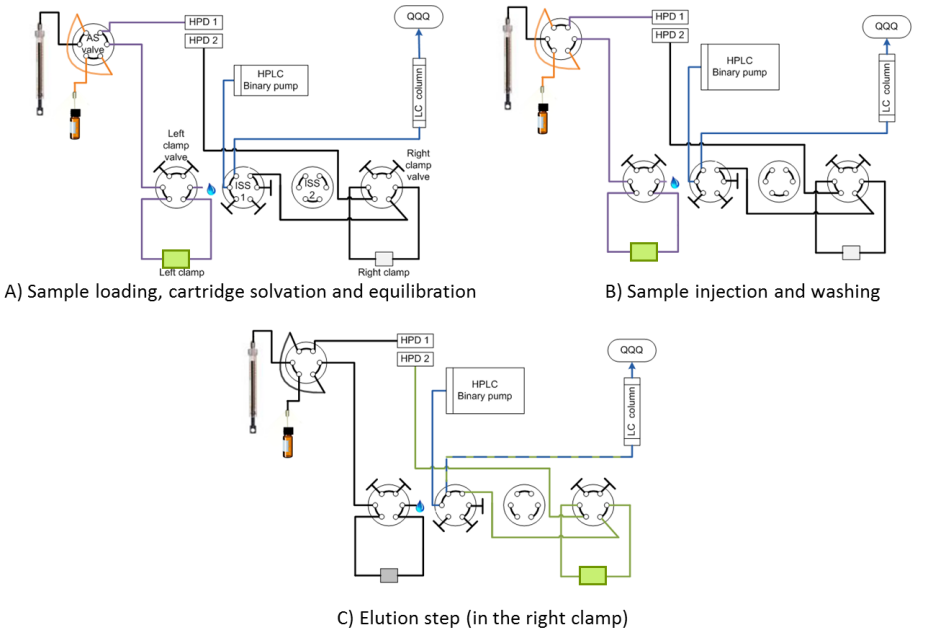


Figure 1. Instrumental configuration including the main steps involved in the SPE step.

This preliminary experiment revealed that the cationic sorbent offered an optimum response for retention/elution of essential amino acids in serum. The acid media favored the retention of the target amino acids, which was complemented with an optimum elution under alkaline conditions to displace the retained analytes from the cationic sorbent.

3.2.2. Optimization of the main variables involved in the SPE process

Once the suitability of the MM cationic sorbent for analysis of essential amino acids in serum was checked, the variables involved in the SPE protocol (*viz.* composition of the solutions and flow-rate) were studied independently for

each SPE step: loading, rinsing and elution. A univariate strategy was used because the discontinuous character of the variables studied. The response factors, aimed at maximizing sensitivity and minimizing dispersion, were the area and height of the chromatographic peaks for the target amino acids.

Methanol and acetonitrile were tested as organic solvents in the loading solution using a dual strategy depending on the solvent concentration. For concentrations of organic solvent below 20% the samples were used as such, whereas superior concentrations (tested up to 100% acetonitrile) required a deproteinization step to avoid tubes clogging by massive precipitation of proteins. Deproteinization was also studied with both organic solvents using different dilution ratios (from 1:1 to 1:10, v/v). The best results were obtained after deproteinization with methanol using a 6:1 (methanol:serum v/v) ratio —600 μL methanol and 100 μL serum sample. By contrast, 90% acetonitrile 1% formic acid was the preferred option to load the sample by enhancing the ionic interaction. The deproteinization and subsequent loading with a high concentration of organic solvent improved the retention of the target analytes, favored the elimination of interferences and matrix effects and enabled to improve the reusability of the SPE cartridge. The pH of the loading solution, 2.1, was slightly below the pKa of the target amino acids except for phenylalanine (pKa=1.8). Therefore, retention of analytes was a mixed contribution of cationic exchange, preferred for all analytes, but also contribution of non-polar/polar interactions ascribed to the polymeric phase.

The rinsing step was also developed with acetonitrile by testing different concentrations. The deproteinization step simplified rinsing of the cartridge by pumping 1 ml 70% acetonitrile prior to elution of the target analytes.

Finally, the elution step was done in the focusing mode with a reduced volume of 90% acetonitrile/5% ammonia and a low flow rate to avoid effect on the chromatographic process developed at acid conditions. The volume of the elution phase, tested from 0.2 to 0.6 ml, provided 0.45 ml as the optimum.

As Table 3 shows, a cleaning sequence formed by four steps was required to prepare the SPE cartridge for the following analysis. A reusability study enabled to report that each cartridge could be reused for three times by setting a cut-off value of 5% in terms of efficiency. The SPE process was completed in 7.4 min, which enables to prepare the following sample while the chromatographic analysis is running. Supplementary Fig. 1 illustrates the TIC and SRM chromatograms obtained after analysis of serum from an atherosclerosis patient.

Table 3. Sequence of operations for the overall SPE process.

Step	**Solvent**	**Volume (ml)**	**Flow rate (ml/min)**	**Time (s)**
Put the selected cartridge into the left clamp				10
Solvation	Acetonitrile	1.00	2.0	36
Equilibration	50% acetonitrile 1% formic acid	1.00	2.0	36
Sample loading	90% acetonitrile 1% formic acid	0.50	2.0	18
Rinsing	70% acetonitrile	1.00	1.0	66
Move cartridge from left to right clamp				5
Elution	90% acetonitrile 5% ammonia	0.45	0.2	135
Move cartridge from right to left clamp				5
Rinsing	Water	1.00	2.0	36
Rinsing	2% formic acid	1.00	2.0	36
Rinsing	Acetonitrile	1.00	2.0	36
Rinsing	Water	1.00	2.0	36
Total time				7 min 35 s

3.3. Validation of the method

The method resulting from optimization of variables was supported on a validation study for characterization in analytical terms. For this purpose, calibration curves were established by dilutions of the stock solution. The correlation coefficients and linear dynamic ranges are shown in Table 4. The sensitivity of the method was evaluated by estimation of the lowest limit of detection (LLOD) and that of quantitation (LLOQ), both determined by injecting in-series dilutions of the multistandard of amino acids to obtain signals 3 and 10 times, respectively, the background noise (average noise value obtained for blank

injections in SRM chromatograms). The absolute concentrations that provided these signals are also listed in Table 4.

Table 4. Analytical figures of merit of the method for analysis of essential amino acids in human serum.

Analyte	LOD (ng on column)	Linear dynamic range (ng on column)		Correlation coefficient
		LOQ	Maximum	
Tryptophan	0.6	1.9	760.0	0.996
Valine	28.0	95.0	760.0	0.969
Leucine/isoleucine	0.6	1.9	760.0	0.996
Methionine	0.6	1.9	760.0	0.993
Threonine	0.6	1.9	760.0	0.996
Phenylalanine	0.6	1.9	760.0	0.996
Lysine	0.6	1.9	100.0	0.995

The accuracy of the method and potential matrix effects were assessed by analysis of non-spiked and spiked serum samples at two concentration levels (100 and 300 ng/mL) and by comparison to the analysis of aqueous solutions of the standard mixture at the two concentrations. The recovery factor for each analyte was studied with the two-cartridge configuration shown in Fig. 2 by the analysis of five replicates of a non-spiked serum pool. Two MM cationic cartridges were in-series located, so after sample injection, the amount of amino acids non retained in the first cartridge was retained in the second. The eluates from both cartridges were sequentially injected into de QqQ analyzer, estimating the concentration retained in each cartridge. The recovery factor in this system was calculated as amount retained in cartridge 1/(amount retained in cartridge 1 + amount retained in cartridge 2). The recoveries ranged between 99.4–99.9% for the target analytes, which ensured quantitative retention/elution of them.

Matrix effects were estimated by analysis of spiked serum and multistandard solutions at the two levels. The peak areas for chromatographic signals obtained from the analysis of the blank (non-spiked serum) were substracted from those reported by the analysis of spiked serum. The recovery factor in this case ranged from 96.9–99.9%, which ensures the absence of matrix

effects. This is justified taking into account that a deproteinization step was previously carried out.

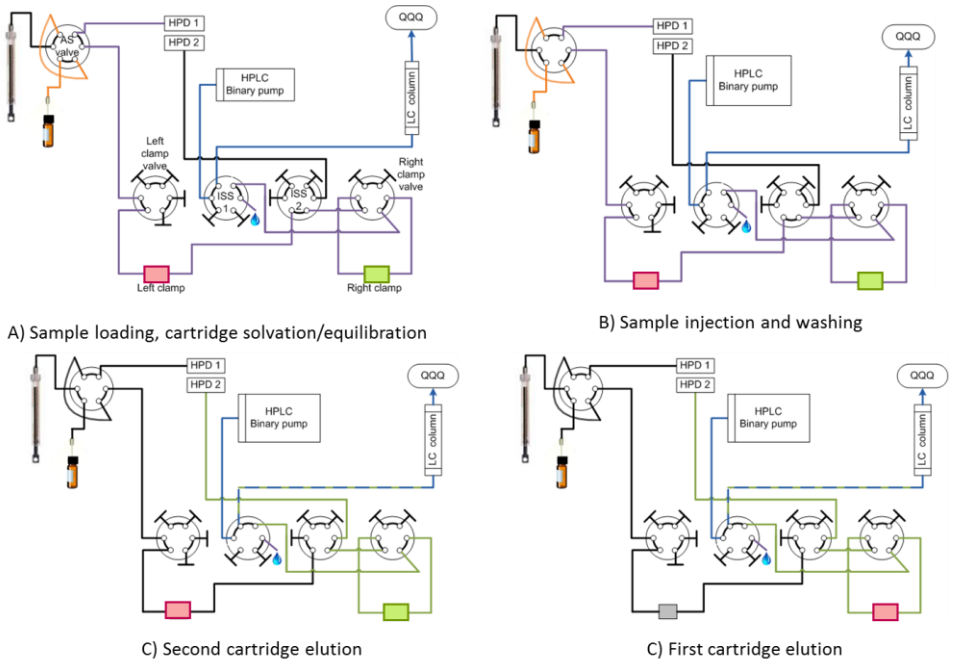


Figure 2. Two-cartridge configuration employed to calculate the recovery factor.

Within-laboratory reproducibility and repeatability were evaluated in a single experimental setup with duplicates by daily experiments carried out for a week with a serum pool . The repeatability, expressed as relative standard deviation (RSD), ranged from 2.3 to 6.0 %, and the within-laboratory reproducibility, also expressed as RSD, ranged from 3.9 to 10.4 %.

3.4. Application of the method to human serum from atherosclerosis patients

Different studies have evidenced that amino acid levels can be related with cardiovascular diseases. The method optimized in this research was applied to human serum sampled from individuals affected by atherosclerosis (n=122), which is one of the most frequent causes of cardiovascular diseases. Eighty of these patients were diagnosed with stable angina while the rest of the individuals

(42) had suffered an acute myocardial infarct. Therefore, the main difference between both groups is that the latter had presented the formation of thrombi with subsequent lesions derived from ischemia.

Table 5. Comparison between normal levels of essential amino acids in plasma and levels obtained in serum of cardiovascular patients.

Analyte	Normal levels (µg/ml)	Levels in cardiovascular patients (µg/ml)
Threonine	16.7±3.9	7.8±3.8
Valine	27.3±5.0	8.0±5.4
Leucine + isoleucine	24.3±5.1	30.9±16.8
Tryptophan	9.0±1.5	7.4±5.0
Lysine	27.5±4.7	5.5±2.0
Methionine	3.7±0.6	24.5±13.1
Phenylalanine	9.4±1.5	45.4±25.6

The concentrations found for each target amino acid in samples from atherosclerosis patients are listed in Table 5. The samples were injected in triplicate resulting in RSD values lower than those estimated in the repeatability test. Averaged levels in healthy individuals for these analytes according to the Human Metabolome Database are also included to compare between them. These results were provided by Cynober after analysis of plasma with the Hitachi analyzer . Anyway, different studies have reported that serum and plasma levels of amino acids are comparable except for dicarboxilic amino acids that are high concentrated in red blood cells . As can be seen, levels of threonine, valine and lysine were below normal levels detected in plasma. On the other hand, leucine/isoleucine, methionine and phenylalanine were found above normal levels in plasma. These results are in partial agreement with other studies found in the literature. Thus, Yao *et al.* have reported elevated levels of threonine and phenylalanine in individuals affected by a myocardial infarction . On the other hand, increased levels of leucine/isoleucine and decreased concentrations of lysine, threonine and phenylalanine were found in plasma from patients who suffered a heart failure as compared to healthy individuals. In the study proposed here, phenylalanine levels were above normal levels in patients with stable angina and with myocardial infarction. On the other hand, threonine levels were below

the normal levels; while a significant increase of threonine level in patients with myocardial infarction as compared to patients with stable angina was found.

Associations between the concentrations of the target amino acids in serum were detected by correlation analysis with a statistical significance at 95% confidence level ($P<0.05$). High correlations were found between all pairs of essential amino acids, except for the tryptophan-phenylalanine and tryptophan-threonine pairs.

The influence of age on amino acids levels was also studied in four groups of patients defined to obtain a normal or Gaussian distribution: 36–50, 51–65, 66–80, and 81–96 years old. With these groups, the skewness and kurtosis coefficients were in the range 0±2 for the studied cohort, which is required for a normal distribution. An ANOVA test revealed no significant differences were detected between groups, so the age does not exert any influence on the concentrations of essential amino acid.

After this characterization, the influence of factors such as obesity, hypercholesterolemia, sex and atherosclerosis on the concentration of essential amino acids was studied. A Multifactor Analysis of Variance (MANOVA) test was applied to detect the influence of these factors and their interactions. This chemometric approach enables to estimate the significance of different sources of variation by analysis of metabolomics data. The variability observed in the data set for each compound is decomposed into contributions to different factors. The contribution of each factor (or interaction between factors) is measured by removing the effects of all other factors and expressing it as a sum of squares. The values and the percent that they represent in the total variability of the compound are shown in Table 6. The concentrations of lysine ($p=0.0335$) and threonine (0.0058) were significant in atherosclerosis patients as a function of either occurrence of ischemia or not. This can be confirmed by visualization of the box-and-whisker plots representing the concentrations of both essential amino acids for patients with stable angina and acute myorcardial infarction (Fig. 3).

Table 6. Contribution of the different factors to amino acids level expressed as sum of squares and the percent each one represents in the total variability observed. The influence of the interaction between pairs of factors is also included.

Factor	Tryptophan	Methonine	Lysine	Threonine	Phenylalanine	Valine	Leucine/Isoleucine
Main effects (Sum of squares (%))							
A: Sex	0.43 ($1{\cdot}10^{-2}$)	112.69 (0.5)	13.22 (2.6)	42.01 (2.4)	947.51 (1.2)	18.82 (0.5)	9.26 ($3{\cdot}10^{-2}$)
B: Cholesterol	0.61 ($2{\cdot}10^{-2}$)	0.01 ($5{\cdot}10^{-5}$)	4.11 (0.8)	6.91 (0.4)	191.80 (0.2)	14.29 (0.4)	89.47 (0.3)
C: Ischemia	0.80 ($3{\cdot}10^{-2}$)	487.35 (2.3)	19.54 (3.8)*	109.00 (6.3)**	49.84 ($6{\cdot}10^{-2}$)	13.44 (0.4)	262.42 (0.8)
D: Obesity	38.20 (1.3)	2540.76 (12.2)***	12.66 (2.4)	43.14 (2.5)	277.57 (0.3)	159.10 (4.5)*	986.58 (2.9)
Interactions (Sum of squares (%))							
AB	3.13 (0.1)	49.97 (0.2)	1.57 (0.3)	16.33 (0.9)	1.39 ($2{\cdot}10^{-3}$)	4.47 (0.1)	75.12 (0.2)
AC	45.77 (1.5)	89.29 (0.4)	30.37 (5.9)**	56.48 (3.2)*	3.25 ($4{\cdot}10^{-3}$)	0.01 ($3{\cdot}10^{-4}$)	8.17 ($3{\cdot}10^{-2}$)
AD	52.68 (1.7)	631.14 (3.0)*	3.16 (0.6)	5.34 (0.3)	313.48 (0.4)	$9{\cdot}10^{-4}$ ($3{\cdot}10^{-5}$)	601.52 (1.8)
BC	39.89 (1.3)	78.91 (0.4)	0.11 ($2{\cdot}10^{-2}$)	3.30 (0.2)	134.69 (0.2)	72.25 (2.1)	15.75 ($5{\cdot}10^{-2}$)
BD	19.61 (0.6)	166.39 (0.8)	0.04 ($8{\cdot}10^{-3}$)	0.93 ($5{\cdot}10^{-2}$)	435.80 (0.5)	146.71 (4.2)*	69.31 (0.2)
CD	2.59 (0.1)	389.74 (1.9)	0.004 ($8{\cdot}10^{-4}$)	2.16 (0.1)	6.82 ($9{\cdot}10^{-3}$)	43.36 (1.2)	130.84 (0.4)

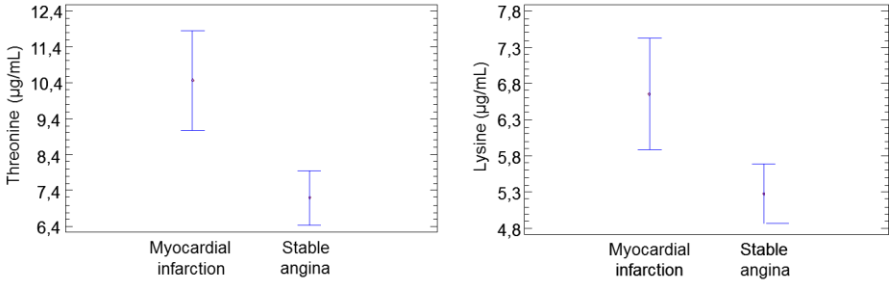


Figure 3. Box-and-whiskers plots for lysine and threonine including mean and 95% least standard deviation for patients with stable angina and myocardial infarction.

As can be seen, serum concentrations of lysine and threonine were higher in patients who had suffered a myocardial infarct. Apart from this statistical significance, the concentration of both analytes was also influenced by the interaction of ischemia–sex factors. Discrimination between male and female individuals is exposed in Fig. 4 that shows a high relevance of sex to explain the variability of the results. Threonine (p=0.0453) and lysine (p=0.0214) levels are considerably increased in those female atherosclerotic patients who had suffered a myocardial infarct. In the case of male individuals, only threonine increased its serum concentration as a result of ischemia. However, lysine level did not report the same behavior since it was slightly decreased due to ischemia.

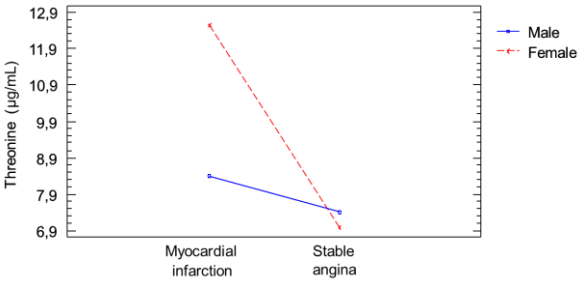


Figure 4. Interaction plot showing the influence of the sex in threonine levels for patients diagnosed with stable angina or myocardial infarction.

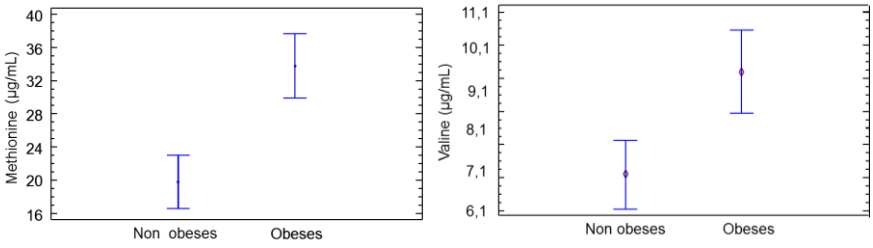


Figure 5. Box-and-whiskers plots for valine and methionine including mean and 95% of least standard deviation for obese and non-obese patients.

Concerning the two risk factors, only obesity contributed to explain the variability in the concentrations of essential amino acids. As can be seen, valine levels (p=0.0167) and, particularly, methionine levels (p=0.0001) were highly influenced by obesity. The box-and-whisker plots, including the 95% least significance difference for the levels of valine and methionine in obese and non-obese atherosclerotic patients is shown in Fig 5. The two plots show significantly increased levels of both amino acids in obese atherosclerotic patients, which could be ascribed to diet.

The interaction between sex and obesity expressed the variation of methionine with statistical significance (p=0.0479). Thus, levels of methionine were higher in serum from non-obese male patients than in non-obese female patiens. The opposite situation was found with obese patients that reported higher methionine concentrations in female atherosclerotic patients (Fig. 6).

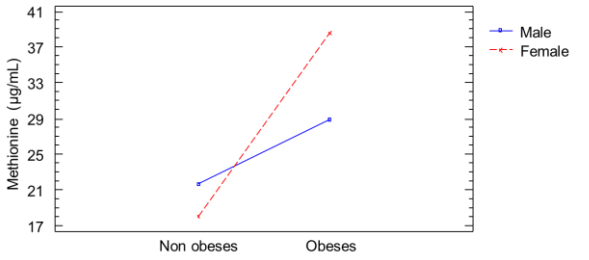


Figure 6. Interaction plot showing the influence of the sex in methionine levels for obese and non-obese patients.

Finally, interaction between obesity and cholesterol explained with statistical significance (p=0.0214) variation on valine levels. In this case (Fig. 7), valine levels were higher in non-obese individuals with hypercholesterolemia and also in obese individuals with normal cholesterol levels.

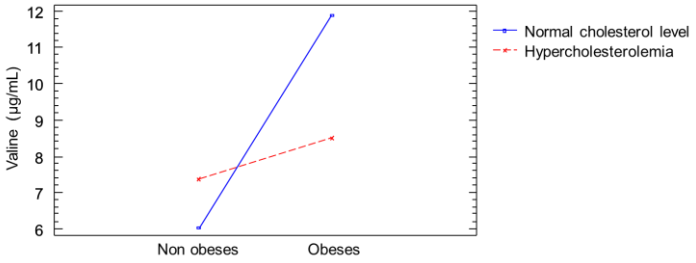


Figure 7. Interaction plot showing the influence of the hypercholesterolemia in valine levels for obese and non-obese patients.

4. Conclusions

An automated method based on an SPE–LC–MS/MS configuration has been developed for analysis of essential amino acids in human serum. The method was optimized for high-throughput analysis of samples with synchronization of the sample preparation and determination steps. The applicability of the method has been supported on analysis of a cohort of atherosclerotic patients diagnosed with stable angina and acute myocardial infarction. The ischemia phenomena, present in patiens who had suffered an infarct, has revealed a significant contribution in the concentration of two essential amino acids such as threonine and lysine. This effect was more pronounced in female than in male individuals. The influence of obesity, as an important risk factor in cardiovascular diseases, has also been elucidated indicating higher levels of valine and methionine in serum from obese individuals.

Acknowledgements

The Spanish Ministerio de Ciencia e Innovación (MICINN) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2009-0734). F. Priego-Capote and M. Calderón Santiago are also grateful to the MICINN for a Ramón y Cajal Contract (RYC-2009-03921) and an FPU scholarship, respectively. The Miguel Servet Hospital (Zaragoza, Spain) is also thanked for donors serum samples.

References

[1] G. Y. Wu, Amino acids: metabolism, functions, and nutrition, Amino Acids 37 (2009).

[2] M. K. Misra, M. Sarwat, P. Bhakuni, R. Tuteja and N. Tuteja, Oxidative stress and ischemic myocardial syndromes, Med. Sci. Monit. 15 (2009).

[3] S. Mitra, T. Goyal and J. L. Mehta, Oxidized LDL, LOX-1 and Atherosclerosis, Cardiovasc. Drugs Ther. 25 (2011).

[4] K. Sugamura and J. F. Keaney, Reactive oxygen species in cardiovascular disease, Free Radic. Biol. Med. 51 (2011).

[5] R. Stocker and J. F. Keaney, Role of oxidative modifications in atherosclerosis, Physiol. Rev. 84 (2004).

[6] N. R. Madamanchi, Z. S. Hakim and M. S. Runge, Oxidative stress in atherogenesis and arterial thrombosis: the disconnect between cellular studies and clinical outcomes, J. Thromb. Haemost. 3 (2005).

[7] M. Elisaf, The treatment of coronary heart disease: An update - Part 1: An overview of the risk factors for cardiovascular disease, Curr. Med. Res. Opin. 17 (2001).

[8] J. O. Hill, J. C. Peters, V. A. Catenacci and H. R. Wyatt, International strategies to address obesity, Obes. Rev. 9 (2008).

[9] P. X. She, C. Van Horn, T. Reid, S. M. Hutson, R. N. Cooney and C. J. Lynch, Obesity-related elevations in plasma leucine are associated with

alterations in enzymes involved in branched-chain amino acid metabolism, Am. J. Physiol.-Endoc. M. 293 (2007).

[10] H. Yao, P. Y. Shi, L. Zhang, X. H. Fan, Q. Shao and Y. Y. Cheng, Untargeted metabolic profiling reveals potential biomarkers in myocardial infarction and its application, Mol. Biosyst. 6 (2010).

[11] G. D. Lewis, R. Wei, E. Liu, E. Yang, X. Shi, M. Martinovic, L. Farrell, A. Asnani, M. Cyrille, A. Ramanathan, O. Shaham, G. Berriz, P. A. Lowry, I. F. Palacios, M. Tasan, F. P. Roth, J. Y. Min, C. Baumgartner, H. Keshishian, T. Addona, V. K. Mootha, A. Rosenzweig, S. A. Carr, M. A. Fifer, M. S. Sabatine and R. E. Gerszten, Metabolite profiling of blood from individuals undergoing planned myocardial infarction reveals early markers of myocardial injury, J. Clin. Inves. 118 (2008).

[12] Y. Li, A. G. Tang and S. Mu, HPLC-FLD determination of serum aromatic amino acids: Application in chronic kidney disease patients, Clin. Chim. Acta 412 (2011).

[13] Y. Y. He, L. J. Zhao, H. Y. Yuan, Z. M. Xu, Y. Tang, D. Xiao and M. M. F. Choi, HPLC with In-Capillary Optical Fiber Laser-Induced Fluorescence Detection of Picomolar Amounts of Amino Acids by Precolumn Fluorescence Derivatization with Fluorescein Isothiocyanate, Chromatographia 74 (2011).

[14] L. J. Song, Z. P. Guo and Y. Chen, One-pot labeling-based capillary zone electrophoresis for separation of amino acid mixture and assay of biofluids, Anal. Chim. Acta 703 (2011).

[15] D. H. Shangguan, Y. X. Zhao, H. W. Han, R. Zhao and G. Q. Liu, Derivatization and fluorescence detection of amino acids and peptides with 9-fluorenylmethyl chloroformate on the surface of a solid adsorbent, Anal. Chem. 73 (2001).

[16] L. Strieglerova, P. Kuban and P. Bocek, Electromembrane extraction of amino acids from body fluids followed by capillary electrophoresis with capacitively coupled contactless conductivity detection, J. Chromatogr. A 1218 (2011).

[17] Y. H. Tak, G. W. Somsen and G. J. de Jong, Optimization of dynamic pH junction for the sensitive determination of amino acids in urine by capillary electrophoresis, Anal. Bioanal. Chem. 401 (2011).

[18] J. LeBoucher, C. Charret, C. CoudrayLucas, J. Giboudeau and L. Cynober, Amino acid determination in biological fluids by automated ion-exchange chromatography: performance of Hitachi L-8500A, Clin. Chem. 43 (1997).

[19] M. T. Ackermans, L. P. Klieverik, P. Ringeling, E. Endert, A. Kalsbeek and E. Fliers, An online solid-phase extraction-liquid chromatography-tandem mass spectrometry method to study the presence of thyronamines in plasma and tissue and their putative conversion from (13)C(6)-thyroxine, J. Endocrinol. 206 (2010).

[20] E. Jagerdeo, M. A. Montgornery, M. Sibum, T. A. Sasaki and M. A. LeBeau, Rapid Analysis of Cocaine and Metabolites in Urine Using a Completely Automated Solid-Phase Extraction-High-Performance Liquid Chromatography-Tandem Mass Spectrometry Method, J. Anal. Toxicol. 32 (2008).

[21] E. Jagerdeo, M. A. Montgomery, M. A. LeBeau and M. Sibum, An automated SPE/LC/MS/MS method for the analysis of cocaine and metabolites in whole blood, J. Chromatogr. B 874 (2008).

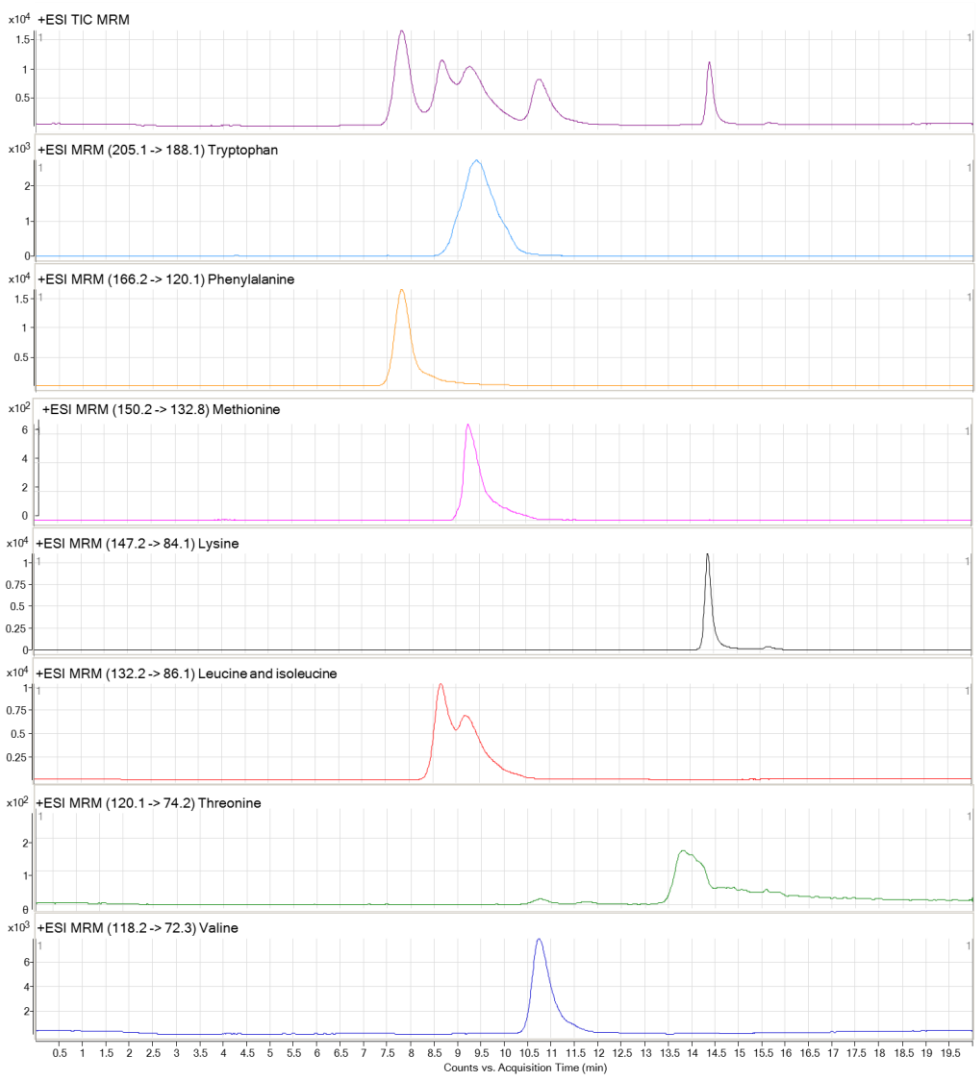
[22] D. L. Massart, B. G. M. Vanderginstre, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, part A, (1997).

[23] L. A. Cynober, Plasma amino acid levels with a note on membrane transport: Characteristics, regulation, and metabolic significance, Nutrition 18 (2002).

[24] J. Wernerman and O. Rooyackers, Methods, who cares?, Clin. Nutr. 19 (2000).

[25] H. Norrelund, H. Wiggers, M. Halbirk, J. Frystyk, A. Flyvbjerg, H. E. Botker, O. Schmitz, J. O. L. Jorgensen, J. S. Christiansen and N. Moller, Abnormalities of whole body protein turnover, muscle metabolism and levels of metabolic hormones in patients with chronic heart failure, J. Intern. Med. 260 (2006).

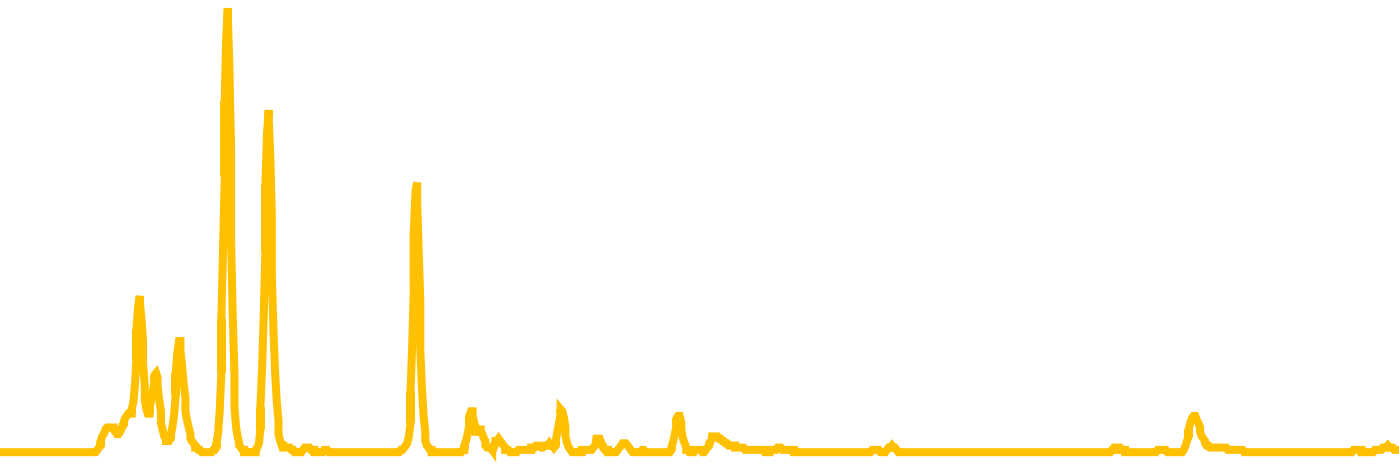
Supplementary material



Supplementary figure 1. TIC and SRM chromatograms obtained after analysis of serum from an atherosclerotic patient.

Chapter 10:

Method based on GC–MS to study the influence of tricarboxylic acids cycle metabolites on cardiovascular risk factors



Journal of Pharmaceutical and Biomedical Analysis

74 (2013) 178–185

Method based on GC–MS to study the influence of tricarboxylic acids cycle metabolites on cardiovascular risk factors

M. Calderón-Santiago [a,b], F. Priego-Capote [a,b], J.G. Galache-Osuna[c], M. D. Luque de Castro[a,b]

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Department of Cardiology, Hemodynamic and Interventional Cardiology, University Hospital Miguel Servet, Zaragoza, Spain*

Method based on GC–MS to study the influence of tricarboxylic acids cycle metabolites on cardiovascular risk factors

*M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M. D. Luque de Castro**

Abstract

Metabolites involved in the tricarboxylic acid (TCA) cycle have previously been proposed as cardiovascular biomarkers. This cycle plays a key role in cell metabolism and the levels of the involved metabolites can also be affected by other physiological factors. The influence of three cardiovascular risk factors such as obesity, hypercholesterolemia, and smoking habit on serum levels of TCA-cycle metabolites has been studied in patients diagnosed with significant coronary lesion. For this purpose, a method based on GC–MS for determination of the target metabolites (*viz.* citric/isocitric, pyruvic, aconitic, oxaloacetic, malic, fumaric and succinic acids) in serum has been developed. The high accuracy and throughput analysis featuring the method have allowed application to a cohort of 223 patients, 172 of them with significant coronary lesion. Multifactor analysis of variance has revealed interactions between the occurrence or not of a coronary lesion and the risk factors considered in this study. These interactions were crucial to explain the levels of target TCA metabolites. Statistical evaluation by ROC curves allowed evaluating the discrimination capability of those significant metabolites with the occurrence of coronary lesions.

1. Introduction

The citric acid cycle, also known as the tricarboxylic acids cycle (TCA cycle) or the Krebs' cycle, consists of a series of chemical reactions aimed at generating energy (ATP) through oxidization of acetate from carbohydrates, fats and proteins into carbon dioxide and water. In addition, the cycle provides precursors for the biosynthesis of compounds including certain amino acids as well as the reducing agent NADH, involved in a number of biochemical reactions (Suppl. Fig. 1). Oxygen is needed along the TCA pathway since the different coenzymes involved in the process need subsequent re-oxidation [1]. Drop of oxygen levels in cells interrupts activity in the Krebs cycle; therefore, the products from glycolysis would not be processed as required.

Alterations in the TCA cycle have been correlated with numerous pathologies such as cancer [2], cardiovascular diseases [3], metabolic syndrome [4] and neurodegenerative disorders [5], where oxidative stress plays a key role [6–9]. The importance of the TCA cycle in the metabolic defensive mechanism against oxidative stress is clear as this cycle is the main producer of pro-oxidant NADH. Thus, alterations in the TCA cycle could modulate the production of reactive oxygen species (ROS) by increasing the concentration of ROS and shortening cellular longevity as a result. The link between TCA cycle and ROS homeostasis may explain the fact that an inefficient functioning of the former characterizes various pathological conditions and ageing [10].

Atherosclerosis, one of the most common cardiovascular diseases, is attributed to the accumulation of lipids, macrophages, smooth muscle cells, etc. in the walls of arteries, thus reducing the lumen of arteries. The result is the appearance of coronary lesions, which make difficult the blood supply of nutrients and oxygen to the heart muscle. The relationship between cardiovascular diseases and the TCA cycle has been supported on the appearance of ROS and oxygen supply deficiency (ischemia) [11]. In fact, metabolites involved in the TCA cycle have been widely suggested as cardiovascular biomarkers [12–15]. Thus, citric acid was found at lower concentration in patients with non-ST elevation acute coronary

syndrome (NSTEACS) as compared to both patients with stable coronary artery disease and controls [12]. Also, succinate and malate were detected at high concentrations in the circulation system of groups affected by myocardial infarction [13], although inhibition of the TCA cycle is supposed on myocardial ischemia. In fact, citric acid level was decreased in groups with clear-cut inducible ischemia [14]. Additionally, α-ketoglutaric acid is within the plasma metabolites that allowed discrimination between individuals with high and low risk of suffering cardiovascular diseases in general terms, according to well-known risk factors such as age, hypertension, hypercholesterolemia, obesity or smoking habits [15]. It can be of interest to know how these factors, together with cardiovascular disorders, can affect the levels of metabolites involved in the TCA cycle.

Different analytical platforms have been developed to determine compounds of the TCA cycle, the most common being GC–MS after derivatization [16–18]. Other strategies such as LC with either fluorescence (after derivatization) or photometry detection [19,20] have also been used. In this research, a method based on GC–MS has been proposed for determination of TCA cycle metabolites in human serum from individuals diagnosed with significant coronary lesions. The objective was to study the potential effects of risk factors for the occurrence of coronary lesions such as obesity, hypercholesterolemia or smoking habit on serum levels of metabolites of the TCA cycle. The data thus obtained have been treated searching for factors that could explain variations in the level of metabolites of the TCA cycle.

2. Material and methods

2.1. Materials, reagents and chemicals

The target metabolites were citric acid, oxaloacetic acid, aconitic acid, isocitric acid, pyruvic acid, α-ketoglutaric acid, malic acid, fumaric acid and succinic acid, supplied by Sigma–Aldrich, Co. (St. Louis, MO, USA). Isotopic (^{15}N) glutamic acid from Sigma–Aldrich was used as internal standard for quantitation of TCA metabolites. Methanol, formic acid and chloroform were

from Panreac (Barcelona, Spain). 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 derivatization solvent. Deionized water (18 MΩ • cm) from a Millipore Milli-Q water purification system was used to prepare the solutions.

2.2. Cohort selected for the study

A total of 223 patients diagnosed with different cardiovascular problems were recruited at the Cardiology Unit of Miguel Servet Hospital (Zaragoza, Spain). Among them, 172 patients were affected by a significant coronary artery lesion as atherosclerosis after evaluation through cardiac catheterization (angiographic stenosis revealed a reduction of the arterial lumen ≥ 70%). Blood samples from these patients were collected within 7 days after catheterism intervention. The rest of the patients (n=51) were affected by other clinical manifestations without significant coronary lesion such as chest pain, angina, and heart failure, or were heart transplanted individuals. The main characteristics of the patients are shown in Table 1 with discrimination between individuals diagnosed with a coronary lesion or not. The average age of the cohort was 66±14, with a similar range for patients with and without coronary lesion. A 70.4% of the total individuals were male, most of them affected by a coronary lesion. Concerning risk factors, 46.2% of total individuals were smokers, 31.4% obese and 45.7% with hypercholesterolemia.

Table 1. Features of the cohort under study.

Characteristic	**Patients (n=223)**	**Patients with coronary lesion (n=172)**	**Patients without coronary lesion (n=51)**	**Fisher's exact test** *p-value*
Age	66 ± 14	67.7 ± 13.4	61.5 ± 12.8	–
Male gender, n (%)	157 (70.4)	136 (79.1)	21 (41.2)	0.000***
Smoking habit, n (%)	103 (46.2)	86 (50.0)	17 (33.3)	0.026*
Obesity, n (%)	70 (31.4)	57 (33.1)	13 (25.5)	0.195
High-cholesterol level, n (%)	102 (45.7)	88 (51.2)	14 (27.5)	0.002**

* P-value between 0.05 and 0.01.
** P-value between 0.01 and 0.001.
*** P-value lower than 0.001.

2.3. Blood extraction and serum isolation

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and incubated for 10 min at room temperature to allow coagulation. Then, the tubes were centrifuged at 2000 × g for 15 min at 4 °C to isolate the serum fraction (processing within 1 h after collection). Serum was placed in plastic ware tubes and stored at –80 °C until analysis.

All steps from blood extraction to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki (2004), and supervised by specialized personnel from Miguel Servet Hospital (Zaragoza, Spain). Individuals selected for this study were previously informed to obtain consent.

2.4. Instruments and apparatus

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 (Afora, Barcelona, Spain) was used to maintain constant the temperature during the derivatization step.

A Varian CP-3800 Gas-Chromatograph coupled to a Saturn 2200 ion-trap mass spectrometer (Sugar Land, TX, USA) equipped with a VF-5ms Factor Four capillary column (30 m x 0.25 mm, 0.25 μm), also provided by Varian, was used for the specific analysis of organic acids in the extracts. The overall system was controlled and the data acquired and processed by using Star Chromatography Workstation 6.0 software (Varian, Walnut Creek, California, USA). All injections were carried out using a Varian 8400 autosampler equipped with a 10 μl syringe from SGE (Scharlab, Barcelona).

2.5. Sample preparation for analysis of tricarboxylic acids

First, 5 μL of the internal standard stock solution at 80 μg/mL was added to the serum sample (150 μL) prior to deproteinization with 300 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 and cleaned by extraction of non-polar compounds with 300 μL of chloroform. The clean phase was concentrated by evaporation and the resulting residue was reconstituted with 10 μL of pyridine and 1 μL of formic acid. For derivatization, 90 μL of a 98:2 BSTFA–TMCS mixture was mixed with the analytical sample, shaken for 30 s and maintained at 30 °C for 30 min before injection into the chromatograph.

2.6. GC–MS analysis

Eight μL of the derivatized solution was injected into the GC instrument with the following split ratio program: 1:50 (0 min), 1:1 (0.5 min), 1:100 (3.5 min) and 1:20 (4.5 min). The injector temperature was programmed as follows: from 70 °C (held 0.5 min) to 270 °C (maintained for 25.5 min) at a rate of 150 °C/min. Carbon dioxide was used as coolant gas for the injector in order to reduce analysis time. The column temperature was programmed from 60 °C (held 1 min) to 300 °C (maintained for 5 min) at a rate of 10 °C/min. Helium was used as carrier gas at a constant flow rate of 1 mL/min.

Electron impact ionization (EI) positive mode was selected prior to single ion monitoring of each analyte by ion trap in microselected ion storage (μSIS) mode. The instrumental parameters were set at the following values: filament emission current 70 μA; transfer line, ion trap and manifold temperatures 280, 200 and 50 °C, respectively. The scan time for data acquisition was set at 1.0 s, and the number of microscans per second for each segment was a function of the mass range under study (always from 50 m/z to ten units above the highest mass monitored in the segment).

2.7. Statistical analysis

Statgraphics Centurion (XV.I version, Statpoint Technologies Inc.), R (URL http://www.R-project.org) and pROC (1.5.1 version, URL http://web.expasy.org/pROC/) were used for data analysis.

The Kolmogorov-Smirnov test was used to check if TCA serum levels showed normal distributions in the cohort selected for the study. Statistical significance analysis (set at 0.05) was carried out by the non-parametric Mann-Whitney test for continuous variables, while the Fisher exact test was used for categorical variables. Multifactor analysis of variance was tested to evaluate the significance of each factor as well as two-factor interactions after normalization of TCA serum levels by box-cox transformation.

A receiver–operating characteristic (ROC) curve was determined in each cohort for each individual predictor, and a cutoff value was selected as the threshold predicting the occurrence of a coronary lession with specificity >90%. Partial ROC AUCs (pAUC) [21,22] were calculated using an adaptation of algorithms previously reported [23]. pAUCs were restricted between 90–100% specificity, considering an efficient predictor in clinical practice should be able to identify clearly at least nine out of ten patients as having a favorable prognosis for negative tests.

3. Results and discussion

3.1. Optimization of sample preparation

Sample preparation was optimized by a serum pool to which the suited volume of internal standard was added to eliminate signal fluctuations due to variations in the injection volume. The protein content in serum justified the inclusion of a deproteinization step, efficiently carried out with methanol in a 2:1 methanol–serum ratio. Methanol has been widely used for protein precipitation by virtue of its high reproducibility and efficiency [24]. After centrifugation, different organic solvents were tested for the removal of organic interferences by

liquid–liquid extraction. An 1:1 methanol–chloroform ratio provided no significant losses of metabolites but a significant decrease of the background signal.

The reaction temperature and time of the derivatization step were also optimized, and the necessity for including an oximation step was tested as some methods for analysis of organic carboxylic acids include this step [18,25,26] while some others do not [16,17,19]. Oximation was found to be unnecessary; therefore, the residue from evaporation was directly reconstituted with 10 μL of pyridine (the optimum medium for derivatization with BSTFA–TMCS). The difficulty for complete dissolution of the residue made necessary the addition of 1 μL of formic acid. After shaking, 90 μL of the derivatization mixture (98:2 BSTFA–TMCS) was added to form trimethylsilyl (TMS) derivatives [27]. Several aliquots of the serum pool were prepared in this way, each maintained at different temperature (from 25 to 60 °C) for different times (from 15 to 60 min) to establish the optimum operational conditions for derivatization of the target metabolites. The best results were obtained at 30 °C for 30 min. Supplementary Fig. 2 shows the effect of temperature and derivatization time on the efficiency of the reaction for three representative metabolites.

3.2. Characterization of the GC–MS method

Isotopic glutamic acid was selected as internal standard to control potential errors in the protocol. The mass spectra of the TMS-derivatized compounds formed in the EI source led to fragment ions predicted from their chemical structure. The monitored ions were in accordance with research in which the analytes were determined using the same derivatizing reagent [16,18]. The quantitation ions for all target compounds, including the internal standard, are shown in Table 2, which also includes the chromatographic retention times. The chromatogram obtained from a serum sample spiked with 100 μg/L of each analyte is shown in Fig. 1.

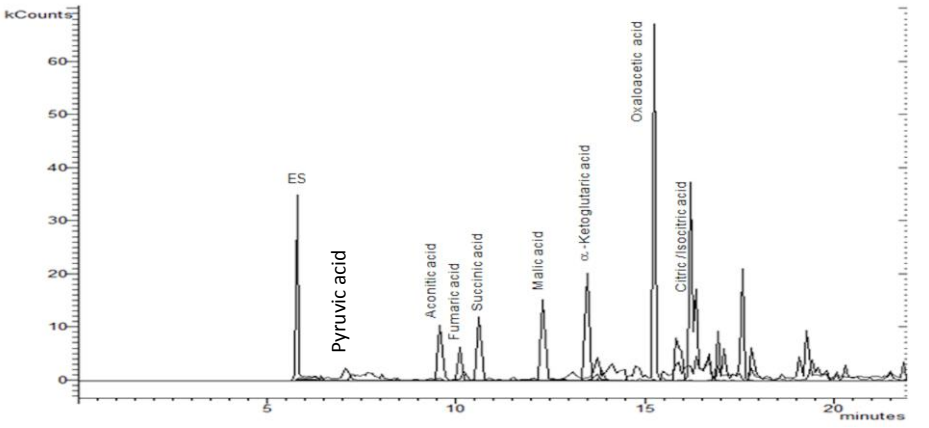


Fig. 1. Single ion monitoring chromatograms from a serum sample spiked with 100 μg/L of the target metabolites and the internal standard (IS).

Table 2. Quantitation ion and retention time for each target compound.

Compound	**Quantitation ion** *(m/z)*	**Retention time** *(min)*
Glutamic acid (isotopic marker)	291.2	5.8
Oxaloacetic acid	290.0	6.1
Pyruvic acid	174.0	6.5
Succinic acid	335.5	9.9
Fumaric acid	278.3	10.4
Malic acid	335.3	12.1
α-Ketoglutaric acid	304.3	13.1
Aconitic acid	229.4	15.1
Citric acid/Isocitric acid	465.4	15.6

The calibration equations were established by using the analyte/internal standard peak-area ratio as a function of the concentration of each analyte in order to circumvent matrix effects. The linear dynamic range for each analyte is shown in Table 3. Characterization of the method also involved calculation of limits of detection (LODs) and quantification (LOQs), which were estimated by diluting a water solution of the target compounds. The LOD for each analyte was expressed as the mass of analyte which gives a signal 3σ above the mean blank

signal (where σ is the standard deviation of the blank signal), and the LOQ as the mass of analyte which gives a signal 10σ above the mean blank signal. Values of LOD and LOQ for each analyte, expressed as concentration in serum, are shown in Table 3. Sensitivity values reported with this method were better than those described in recently published methods, which presented minimum LOD values of 10 and 20 μg/L for α-ketoglutaric acid and citric acid, respectively [29], while in other cases LODs from 56 to 140 μg/L [24,28] have been reported.

Table 3. Limits of detection (LOD) and quantitation (LOQ) for each target compound and linear dynamic range of the calibration curve.

Compound	**LOD** (μg/L)	**Linear dynamic range** (μg/L)	
		LOQ	**Maximum limit**
Oxaloacetic acid	2.00	6.67	388
Pyruvic acid	10.00	33.33	317
Succinic acid	2.00	6.67	1107
Fumaric acid	2.00	6.67	626
Malic acid	2.00	6.67	526
α-Ketoglutaric acid	10.00	33.33	425
Aconitic acid	2.00	6.67	321
Citric/Isocitric acid	20.00	66.70	402

The precision of the method was calculated by injecting different aliquots of the same sample twice a day for seven days. The results obtained for each compound in terms of intra-day and inter-days precisions, expressed as percent of relative standard deviation (RSD), are shown in Table 4. Those values were quite acceptable for analysis of biological samples. As an example, Lu *et al.* [29] obtained intraday variability values twice or six times higher than those obtained with our method. The accuracy of the method was calculated by comparing a serum sample spiked with a known quantity of analyte with the same serum sample without spiking. Results obtained are shown in Table 4.

Table 4. Precision study expressed as percent of relative standard deviation (% RSD) and accuracy obtained expressed as recovery percentage.

Compound	Intra-day precision	Inter-day precision	Accuracy
Oxaloacetic acid	11.4	17.4	102.7
Pyruvic acid	7.7	17.2	100.8
Succinic acid	5.3	15.8	99.8
Fumaric acid	5.2	5.8	100.1
Malic acid	4.6	6.4	100.0
α-Ketoglutaric acid	7.7	14.8	101.9
Aconitic acid	3.2	7.9	109.5
Citric/Isocitric acid	9.4	15.4	95.4

3.3. Application of the method to human serum samples from the cohort under study

As commented under "Introduction", different studies have evidenced that the levels of compounds involved in the TCA cycle could be related to cardiovascular diseases. In this study, the relationship between TCA serum levels and the diagnosis of coronary lesions has been studied by considering different risk factors influencing the occurrence of cardiovascular disorders. The risk factors studied were obesity, hypercholesterolemia and smoking habit. Apart from them, an anthropometric factor such as individual gender was considered in the statistical analysis. The study of categorical variables revealed statistical significance of three factors by application of the Fisher's exact test (see Table 1): smoking habit (p = 0.026) and, specially, hypercholesterolemia (p = 0.002) and gender (p = 0.000). The obesity did not report a statistical contribution to the presence of coronary lesion in the target cohort.

For analysis of TCA in serum, the samples were injected in triplicate, and the RSD values were of the same order as those estimated in the precision study. Average concentration for each target compound, with the corresponding standard deviation, differentiated patients with significant coronary lesion from

those without coronary lesion, as shows Table 5. Since TCA concentrations did not show normal distributions (Kolmogorov–Smirnov test) between-group differences were tested with the non-parametric Mann–Whitney U test. This non-parametric test did not detect significance of any metabolite between individuals with or without coronary lesion. Prior to application of this test, the influence of individuals age on the levels of target metabolites was studied by grouping the patients to obtain a normal or Gaussian distribution: 36–50, 51–65, 66–80, and 81–96 year old. With these ranges, the skewness and kurtosis standard were between 0±2, fulfilling the condition required for normal distributions. No significant differences were detected between groups, so the age did not influence levels of the target compounds.

Table 5. Levels of TCA cycle metabolites in the cohort under study expressed as mean±standard deviation (μg/L).

Compound	**With significant coronary lesion** *Mean ± SD (n=3)*	**Without coronary lesion** *Mean ± SD (n=3)*	**Mann-Whitney U test** *P-value*
Oxaloacetic acid	50.9 ± 28.9	49.0 ± 20.6	0.3903
Pyruvic acid	152.8 ± 74.2	134.3 ± 88.1	0.1227
Succinic acid	151.2 ± 76.7	219.3 ± 126.9	0.5411
Fumaric acid	164.9 ± 27.0	164.9 ± 23.7	0.6070
Malic acid	43.3 ± 27.5	52.1 ± 33.9	0.2028
α-Ketoglutaric acid	64.3 ± 57.4	80.9 ± 57.9	0.2565
Aconitic acid	150.4 ± 77.3	175.7 ± 108.9	0.9169
Citric/Isocitric acid	228.0 ± 147.5	238.6 ± 160.1	0.6297

Interactions between categorical and quantitative variables were analyzed by a multifactor analysis of variance. The variability of each compound was decomposed into contributions from different factors. The contribution of each factor (or interaction between factors) was measured by removing the effects of all other factors and expressing the contribution as a sum of squares. The values and the percent they represented in the total variability of the compounds are shown in Suppl. Table 1. As can be seen, risk factors were very

influential on the levels of pyruvic acid, oxaloacetic acid, aconitic acid and fumaric acid. In particular, the smoking habit contributed to explain the levels of pyruvic acid (p = 0.0456), oxaloacetic acid (p = 0.0141) and fumaric acid (p = 0.0254); therefore, smoking is the most influential of the risk factors under study on the level of these TCA. Concerning aconitic acid, its serum level was significantly explained by the occurrence of coronary lesions. In the case of pyruvic acid, Fig. 2A shows that its serum level was higher in smoking patients, which is justified by the level of oxygen in blood from smokers, which is lower than in non-smokers and explains the decreased efficiency of the Krebs cycle in the latter.

Oxalacetic and fumaric acids, integrated in the last pathways of the Krebs cycle (see Fig. 2A), showed a behavior similar to that of pyruvic acid. Concerning interaction coronary lesion–aconitic acid, Fig. 2B shows that the concentration of this TCA was lower in individuals diagnosed with coronary lesion than in the rest of individuals in the cohort.

No significant contribution was found for the resting factors under study, namely, the individuals gender, obesity or hypercholesterolemia, although the interaction effect between factors was critical. Thus, the analysis of two-factor interactions revealed the importance of certain risk factors to explain serum levels of the target TCA metabolites. Two of the main interactions were coronary lesion × smoking habit and coronary lesion × obesity, with highly-significant influence on pyruvic acid (p = 0.0029) as well as malic acid (p = 0.0214) and α-ketoglutaric acid (p = 0.0021). As Fig. 3 shows, the influence of obesity and smoking habit was critical to understand these results since an opposite effect was observed depending on obese/non-obese and smoker/non-smoker conditions of the individuals. Concerning α-ketoglutaric and malic acids, serum levels were considerably higher in obese individuals diagnosed with a coronary lesion, but not in non-obese individuals. Pyruvic acid also revealed a contradictory behavior since smoker individuals diagnosed with coronary lesion reported higher concentrations than smoking individuals without coronary lesions and also than non-smoking individuals (Fig. 3B). One other interaction

contributing to explain levels of TCA metabolites was the gender × smoking habit, which influenced serum concentrations of pyruvic and succinic acids. The behavior observed for both TCA metabolites is shown in Suppl. Fig. 3. Previous studies have reported that gender can also affect metabolites involved in the TCA cycle since men generally burn calories more quickly than women as the former posses more muscle tissue [30]. This fact could explain the levels found in non-smoking individuals, but not the change observed for smokers, which preferentially affects women. This may be due to the fact that women are more vulnerable than men to tobacco smoking effects, shown by more breathlessness, coughing and bronquitis [31].

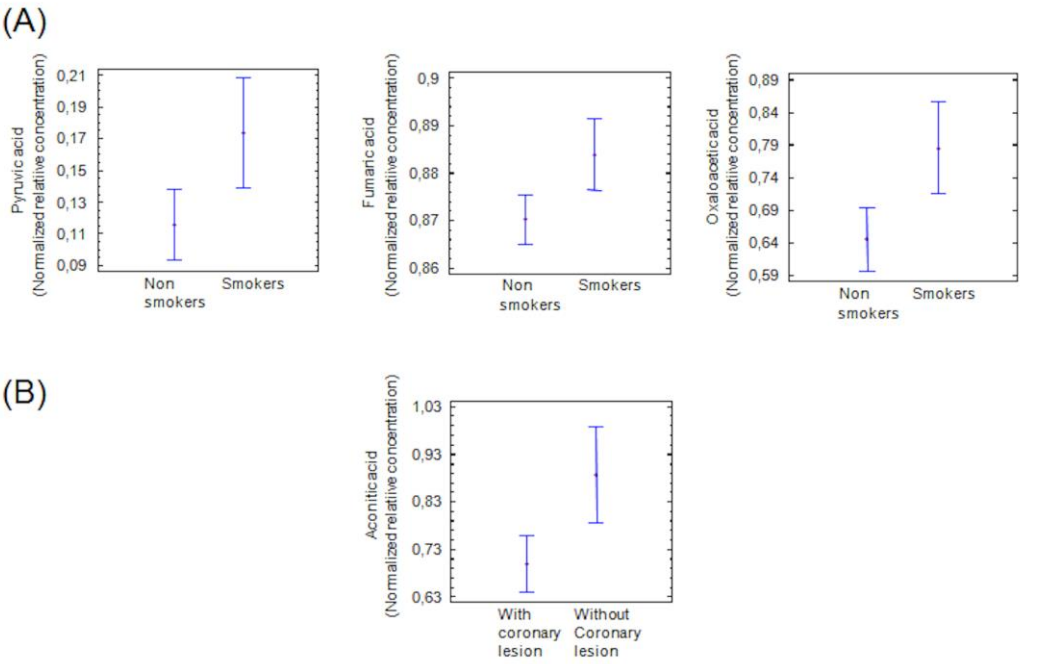


Fig. 2. Mean normalized concentrations in serum (95% of standard deviation) for **(A)** pyruvic, fumaric and oxaloacetic acids in patients with/without smoking habit; **(B)** aconitic acid in patients with/without coronary lesion. The concentration values were normalized by dividing each value by the highest concentration value.

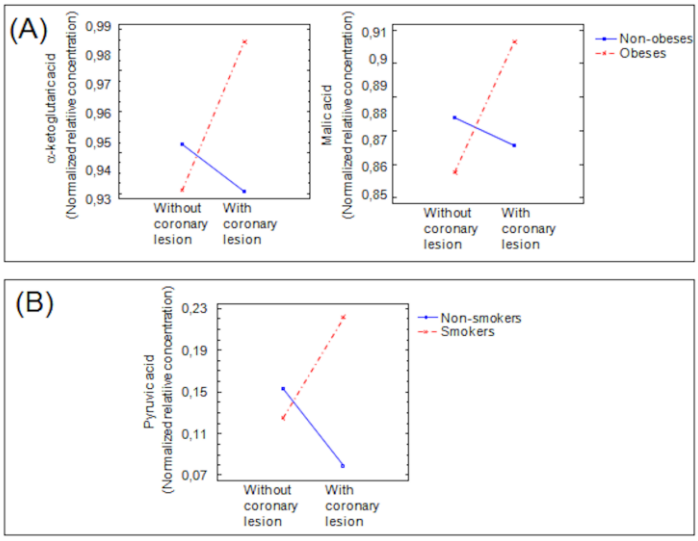


Fig. 3. Interaction plot of: **(A)** obesity × significant coronary lesion on α-ketoglutaric and malic acid levels; **(B)** smoking habit × significant coronary lesion on pyruvic acid levels.

3.4. Evaluation of the TCA predictors by ROC curves

The multifactor analysis of variance revealed the influence of coronary lesions on the variability of concentrations of TCA metabolites in serum, directly (aconitic acid) or by interaction with risk factors such as smoking habit (pyruvic acid) or obesity (malic and α-ketoglutaric acids). These results were validated by ROC curves and by estimation of the pAUC. Evaluation of a predictor by the total AUC is not recommended when the performance test only takes place in high specificity or high sensitivity regions [32]. For this reason, the pAUC parameter is more suited since it is restricted to specific regions of the curve. In this study, aconitic acid was evaluated by one-factor ROC curves (ocurrence or not of coronary lesion), while malic-, α-ketoglutaric- and pyruvic acids were evaluated by a two-factor ROC curves to take into account the interaction effects detected in the multifactor analysis of variance. Thus, malic and α-ketoglutaric acids were evaluated by the occurrence of coronary lesion and obesity as factors, while pyruvic acid was evaluated by the occurrence of coronary lesion and smoking habit as factors. Figure 4 shows the ROC curves for the four TCA metabolites

highlighting the specificity threshold (above 90%) for each TCA metabolite to reduce the probability of false negative predictions. According to the ROC curves, aconitic and pyruvic acids were able to discriminate patients with or without coronary lesions (with smoking habit or not for pyruvic acid) with high specificity (92 and 91%, respectively), but with low sensitivity (21 and 17%, respectively). The situation was different for the other two TCA metabolites the concentration profiles of which were significantly explained by the interaction between coronary lesion and obesity. Thus, malic acid was able to discriminate individuals with coronary lesions with specificity of 92% and sensitivity of 49%, while α-ketoglutaric acid reported values of 93% specificity and 56% sensitivity. The α-ketoglutaric acid seems to have a higher discrimination capability than malic acid. However, these results need to be supported on the pAUC parameter of the two ROC curves, which was 2.3% for α-ketoglutaric acid and 4.2% for malic acid. These values enable to conclude that α-ketoglutaric acid discrimination capability is less robust than that observed for malic acid.

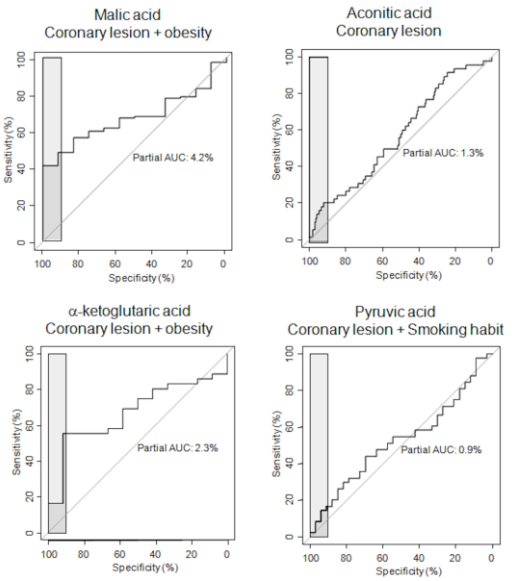


Fig. 4. ROC curves of malic, aconitic, α-ketoglutaric and pyruvic acids for the one- factor or two- factors combination for which each acid was significant. The vertical light grey shape corresponds to the pAUC region (between 90 and 100% specificity).

4. Conclusions

The relevance of three risk factors such as hypercholesterolemia, obesity and smoking habit on cardiovascular diseases has been studied through their influence on serum levels of TCA metabolites for patients diagnosed with coronary lesions. The consideration of these risk factors was crucial to explain the variability of the target TCA for patients with coronary lesion *versus* control individuals. TCA metabolites are primary metabolites that could be affected by numerous internal and external aspects. For this reason, these factors should be considered when TCA cycle metabolites are proposed as potential biomarkers in clinical diagnosis. In this study, the statistical analyses were supported on ROC curves and their related parameters to validate the discrimination capability of TCA metabolites. Statistical analysis revealed that malic acid and α-ketoglutaric acid led to models with high specificity and acceptable sensitivity for discrimination between patients with and without coronary lesions. The discrimination model of α-ketoglutaric acid was less robust than that of malic acid.

Acknowledgements

The Spanish Ministerio de Ciencia e Innovación (MICINN) and the FEDER Program are gratefully acknowledged for financial support (project No CTQ2009-7430). F. Priego-Capote and M. Calderón Santiago are also grateful to the MICINN for a Ramón y Cajal Contract (RYC-2009-03921) and an FPU scholarship, respectively. The Miguel Servet Hospital (Zaragoza, Spain) is also thanked for donors' serum samples.

References

[1] S.L. Seager, M.R. Slabaugh, Organic and Biochemistry for Today, seventh ed., Brooks/Cole, Belmont, 2010.

[2] A.P. John, Dysfunctional mitochondria, not oxygen insufficiency, cause cancer cells to produce inordinate amounts of lactic acid: the impact of this on the treatment of cancer, Med. Hypotheses 57 (2001) 429–431.

[3] M.J. Gibala, M.E. Young, H. Taegtmeyer, Anaplerosis of the citric acid cycle: role in energy metabolism of heart and skeletal muscle, Acta Physiol. Scand. 168 (2000) 657–665.

[4] H.J. Atherton, N.J. Bailey, W. Zhang, J. Taylor, H. Major, J. Shockcor, K. Clarke, J.L. Griffin, A combined H-1-NMR spectroscopy- and mass spectrometry-based metabolomic study of the PPAR-alpha null mutant mouse defines profound systemic changes in metabolism linked to the metabolic syndrome, Physiol. Genomics 27 (2006) 178–186.

[5] X.W. Zhu, G. Perry, P.I. Moreira, G. Aliev, A.D. Cash, K. Hirai, M.A. Smith, Mitochondrial abnormalities and oxidative imbalance in Alzheimer disease, J. Alzheimer's Dis. 9 (2006) 147–153.

[6] B. Hansel, P. Giral, E. Nobecourt, S. Chantepie, E. Bruckert, M.U. Chapman, A. Kontush, Metabolic syndrome is associated with elevated oxidative stress and dysfunctional dense high-density lipoprotein particles displaying impaired antioxidative activity, J. Clin. Endocr. Metab. 89 (2004) 4963–4971.

[7] N.S. Dhalla, R.M. Temsah, T. Netticadan, Role of oxidative stress in cardiovascular diseases, J. Hypertens. 18 (2000) 655–673.

[8] B. Uttara, A.V. Singh, P. Zamboni, R.T. Mahajan, Oxidative Stress and Neurodegenerative Diseases: A Review of Upstream and Downstream Antioxidant Therapeutic Options, Curr. Neuropharmacol. 7 (2009) 65–74.

[9] C. Echiburu-Chau, D. Roy, G.M. Calaf, Metastatic suppressor CD44 is related with oxidative stress in breast cancer cell lines, Int. J. Oncol. 39 (2011) 1481–1489.

[10] R.J. Mailloux, R. Beriault, J. Lemire, R. Singh, D.R. Chénier, R.D. Hamel, V.D. Appanna, The Tricarboxylic Acid Cycle, an Ancient Metabolic Network with a Novel Twist, PloS ONE 2(8) (2007) e690.

[11] S. Muravchick, Clinical implications of mitochondrial disease, Adv. Drug Deliver. Rev. 60 (2008) 1553–1560.

[12] M. Vallejo, A. Garcia, J. Tunon, D. García-Martínez, S. Angulo, J.L. Martin-Ventura, L.M. Blanco-Colio, P. Almeuda, J. Egido, C. Barbas, Plasma fingerprinting with GC-MS in acute coronary syndrome, Anal. Bioanal. Chem. 394 (2009) 1517–1524.

[13] H. Yao, P.Y. Shi, L. Zhang, X.H. Fan, Q. Shao, Y.Y. Cheng, Untargeted metabolic profiling reveals potential biomarkers in myocardial infarction and its application, Mol. Biosyst. 6 (2010) 1061–1070.

[14] M.S. Sabatine, E. Liu, D.A. Morrow, E. Heller, R. McCarroll, R. Wiegand, G.F. Berriz, F.P. Roth, R.E. Gerszten, Metabolomic identification of novel biomarkers of myocardial ischemia, Circulation 112 (2005) 3868–3875.

[15] P. Bernini, I. Bertini, C. Luchinat, L. Tenori, A. Tognaccini, The Cardiovascular Risk of Healthy Individuals Studied by NMR Metabonomics of Plasma Samples, J. Proteome Res. 10 (2011) 4983–4992.

[16] A. Ghassempour, S. Nojavan, Z. Talebpour, A.A. Amiri, N.M. Najafi, Monitoring of the fermentation media of citric acid by the trimethylsilyl derivatives of the organic acids formed, J. Agric. Food Chem. 52 (2004) 6384–6388.

[17] Rosenqvi.H, H. Kallio, V. Nurmiko, Gas-chromatographic analysis of citric-acid cycle and related compounds from Escherichia-Coli as their trimethylsilyl derivatives, Anal. Biochem. 46 (1972) 224–231.

[18] R.S. Kombu, H. Brunengraber, M.A. Puchowicz, Analysis of the Citric Acid Cycle Intermediates Using Gas Chromatography-Mass Spectrometry. In: Metz TO, editor. Metabolic Profiling. Methods and Protocols. Humana Press (Springer), New York, 2011, 147–157.

[19] S. Paavilainen, T. Korpela, Comparison of High-Performance Liquid and Gas-Chromatography in the determination of organic-acids in culture media of Alkaliphilic Bacteria, J. Chromatogr. 634 (1993) 273–280.

[20] K. Kubota, T. Fukushima, R. Yuji, H. Miyano, K. Hirayama, T. Santa, K. Imai, Development of an HPLC-fluorescence determination method for carboxylic acids related to the tricarboxylic acid cycle as a metabolome tool, Biomed. Chromatogr. 19 (2005) 788–795.

[21] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.

[22] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinformatics 12 (2011) 77–84.

[23] J. Carpenter, J. Bithell, Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, Stat. Med. 19 (2000) 1141–1164.

[24] B.M. Wagner, F. Donnarumma, R. Wintersteiger, W. Windischhofer, H.J. Leis, Simultaneous quantitative determination of alpha-ketoglutaric acid and 5-hydroxymethylfurfural in human plasma by gas chromatography-mass spectrometry, Anal. Bioanal. Chem. 396 (2010) 2629–2637.

[25] J. A, 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.

[26] L. Yang, T. Kasumov, L. Yu, K.A. Jobbins, F. David, S.F. Previs, J.K. Kelleher, H. Brunengraber, Metabolomic assays of the concentration and mass isotopomer distribution of gluconeogenic and citric acid cycle intermediates, Metabolomics 2 (2006) 85–94.

[27] D.R. Knapp, Handbook of analytical derivatization reactions, first ed., Wiley-Interscience, New York, 1979.

[28] D. Kloos, R.J.E. Derks, M. Wijtmans, H. Lingeman, O.A. Mayboroda, A.M. Deelder, W.M.A. Niessen, M. Giera, Derivatization of the tricarboxilic acid cycle intermediates and analysis by online solid-phase extraction-liquid chromatography-mass spectrometry with positive-ion electrospray ionization, J. Chrom. A 1232 (2012) 19–26.

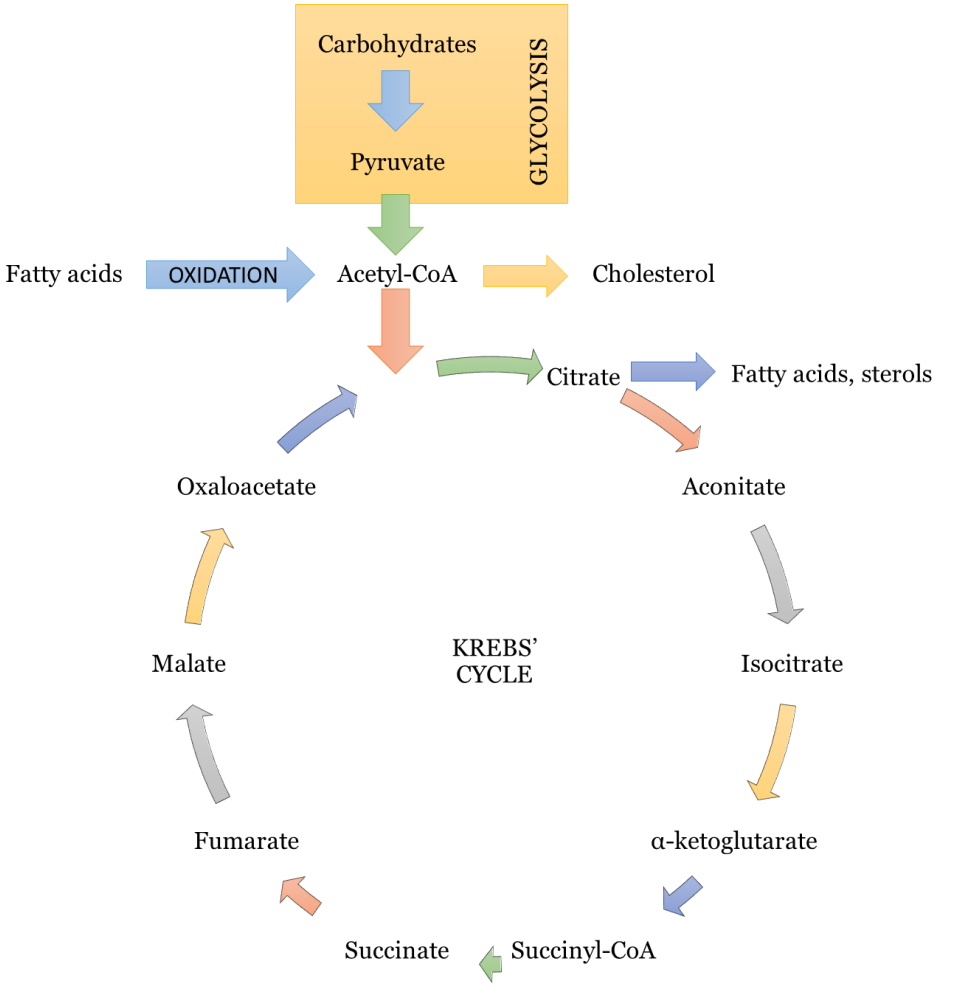
[29] S. Lu, X. Sun, C. Shi, Y. Zhang, Determination of tricarboxilic acid cycle acids and other related substances in cultured mammalian cells by gradient ion-exchange chromatography with suppressed conductivity detection, J. Chrom. A 1012(2) (2003) 161–168.

[30] H.J. Green, I.G. Fraser, D.A. Ranney, Male and female differences in enzyme-activities of energy-metabolism in vastus lateralis muscle, J. Neurol. Sci. 65 (1984) 323–331.

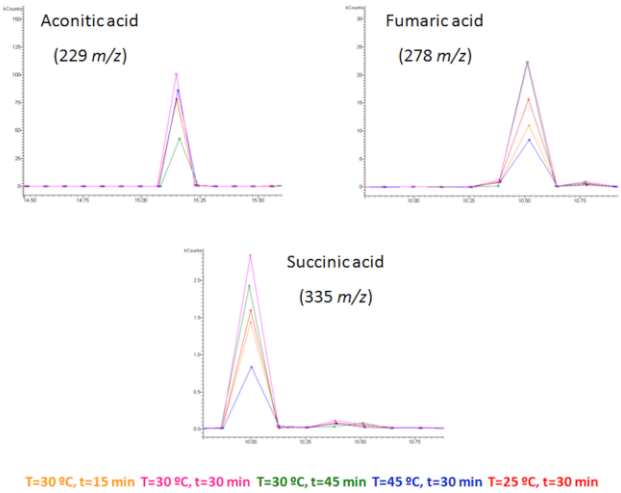
[31] A. Langhammer, R. Johnsen, J. Holmen, A. Gulsvik, L. Bjermer, Cigarette smoking gives more respiratory symptoms among women than among men - The Nord-Trondelag Health Study (HUNT), J. Epidemiol. Commun. H. 54 (2000) 917–922.

[32] G. Stranjalis, S. Korfias, C. Psachoulia, A. Kouyialis, D.E. Sakas, A.D. Mendelow, The prognostic value of serum S-100B protein in spontaneous subarachnoid haemorrhage, Acta Neurochir. 149 (2007) 231–238.

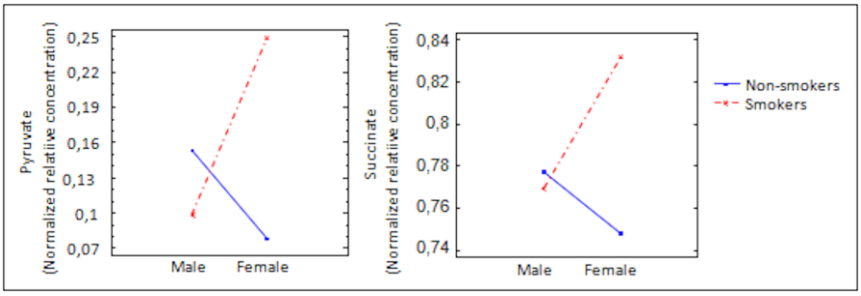
Supplementary material



Supplementary Fig. 1. Krebs' cycle scheme including interactions with cholesterol, fatty acids and carbohydrates.



Supplementary Fig. 2. GC–MS peaks of aconitic, fumaric and succinic acids obtained at different derivatization times and temperatures in the optimization step.



Supplementary Fig. 3. Interaction plot of smoking habit × gender on pyruvic and succinic acid levels.

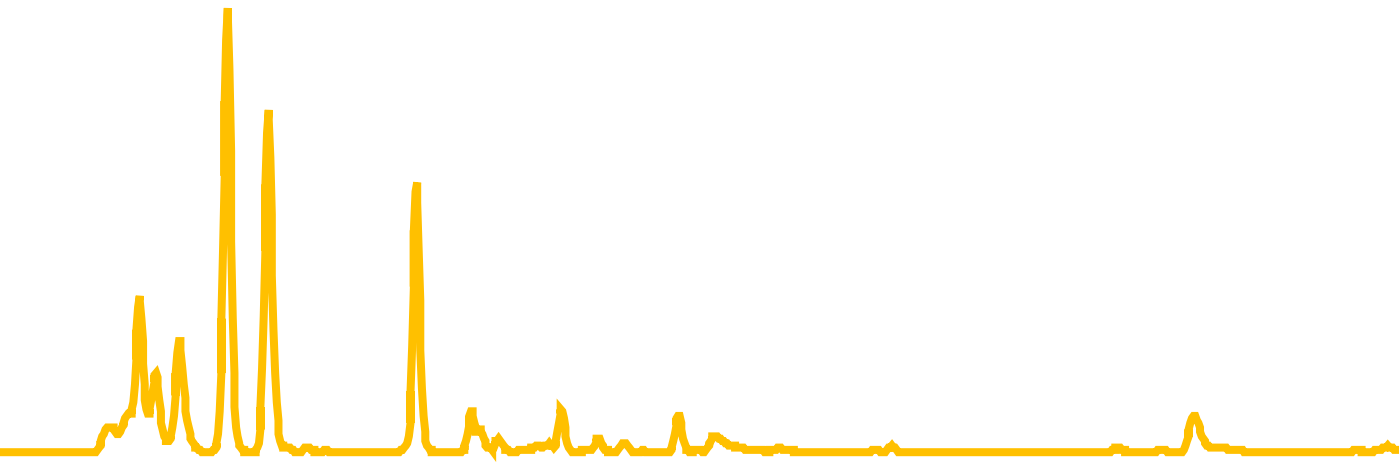
Supplementary Table 1. Results from the multifactor analysis of variance, including the variability of each target compound decomposed into contributions from different factors and their interactions. All contributions are expressed as a sum of squares, including the percent of their contribution to the total variability for each compound.

	Oxaloacetic acid	**Pyruvic acid**	**Succinic acid**	**Fumaric acid**	**Malic acid**	**α-Ketoglutaric acid**	**Aconitic acid**	**Citric/Isocitric acid**
Main effects (Sum of squares (%))								
A: Gender	0.23 (1.46)	$2.87\cdot10^{-2}$ (0.73)	$5.54\cdot10^{-3}$ (0.28)	$5.69\cdot10^{-4}$ (0.32)	$2.47\cdot10^{-3}$ (0.48)	$1.06\cdot10^{-3}$ (0.26)	$1.83\cdot10^{-2}$ (0.07)	$6.02\cdot10^{-3}$ (0.05)
B: Hypercholesterolemia	$6.56\cdot10^{-2}$ (0.42)	$3.99\cdot10^{-2}$ (1.01)	$1.37\cdot10^{-2}$ (0.68)	$1.19\cdot10^{-3}$ (0.67)	$3.52\cdot10^{-4}$ (0.07)	$2.81\cdot10^{-3}$ (0.70)	0.31 (1.16)	0.12 (0.90)
C: Coronary lesion	$1.94\cdot10^{-2}$ (0.12)	$8.09\cdot10^{-3}$ (0.20)	$1.6\cdot10^{-2}$ (0.80)	$3.46\cdot10^{-4}$ (0.19)	$6.80\cdot10^{-3}$ (1.31)	$5.71\cdot10^{-3}$ (1.43)	0.65 (2.46)*	0.15 (1.15)
D: Smoking	0.44 (2.82)*	$6.74\cdot10^{-2}$ (1.71)*	$3.37\cdot10^{-2}$ (1.68)	$4.14\cdot10^{-3}$ (2.31)*	$6.01\cdot10^{-3}$ (1.16)	$4.19\cdot10^{-3}$ (1.05)	0.21 (0.81)	0.22 (1.70)
E: Obesity	0.17 (1.09)	$5.94\cdot10^{-2}$ (1.51)	$7.45\cdot10^{-4}$ (0.04)	$5.36\cdot10^{-4}$ (0.30)	$1.37\cdot10^{-3}$ (0.26)	$5.57\cdot10^{-3}$ (1.40)	$4.85\cdot10^{-3}$ (0.02)	$5.32\cdot10^{-3}$ (0.04)
Interactions (Sum of squares (%))								
AB	0.16 (1.00)	$2.11\cdot10^{-2}$ (0.54)	$6.62\cdot10^{-3}$ (0.33)	$1.01\cdot10^{-3}$ (0.56)	$7.42\cdot10^{-4}$ (0.14)	$1.04\cdot10^{-3}$ (0.26)	0.14 (0.51)	$2.23\cdot10^{-3}$ (0.02)
AC	$5.16\cdot10^{-2}$ (0.33)	$5.48\cdot10^{-2}$ (1.39)	$2.89\cdot10^{-3}$ (0.14)	$1.84\cdot10^{-4}$ (0.10)	$2.37\cdot10^{-5}$ (0.005)	$3.58\cdot10^{-5}$ (0.01)	$4.22\cdot10^{-3}$ (0.02)	$2.52\cdot10^{-2}$ (0.19)
AD	0.18 (1.19)	0.25 (6.39)***	$4.89\cdot10^{-2}$ (2.44)*	$2.12\cdot10^{-3}$ (1.18)	$5.57\cdot10^{-3}$ (1.08)	$6.31\cdot10^{-3}$ (1.58)	$7.46\cdot10^{-2}$ (0.28)	0.19 (1.45)
AE	$8.71\cdot10^{-2}$ (0.56)	$4.81\cdot10^{-2}$ (1.22)	$3.34\cdot10^{-2}$ (1.67)	$1.12\cdot10^{-3}$ (0.63)	$7.92\cdot10^{-3}$ (1.53)	$4.23\cdot10^{-3}$ (1.06)	0.19 (0.73)	0.13 (0.99)
BC	0.13 (0.81)	$2.28\cdot10^{-4}$ (0.006)	$7.83\cdot10^{-3}$ (0.39)	$1.67\cdot10^{-4}$ (0.09)	$2.00\cdot10^{-3}$ (0.39)	$5.80\cdot10^{-3}$ (1.45)	$3.85\cdot10^{-2}$ (0.15)	$6.48\cdot10^{-2}$ (0.50)
BD	0.11 (0.70)	$6.87\cdot10^{-3}$ (0.17)	$1.69\cdot10^{-2}$ (0.84)	$8.24\cdot10^{-4}$ (0.46)	$1.57\cdot10^{-4}$ (0.03)	$2.20\cdot10^{-4}$ (0.06)	0.10 (0.39)	$2.22\cdot10^{-2}$ (0.17)
BE	$8.43\cdot10^{-4}$ (0.005)	$2.74\cdot10^{-4}$ (0.007)	$2.41\cdot10^{-3}$ (0.12)	$6.70\cdot10^{-4}$ (0.37)	$6.95\cdot10^{-5}$ (0.01)	$1.74\cdot10^{-4}$ (0.04)	$1.44\cdot10^{-2}$ (0.05)	$1.18\cdot10^{-2}$ (0.09)
CD	$3.97\cdot10^{-3}$ (0.03)	0.15 (3.85)**	$1.4\cdot10^{-2}$ (0.70)	$7.16\cdot10^{-4}$ (0.40)	$1.08\cdot10^{-3}$ (0.21)	$9.20\cdot10^{-4}$ (0.23)	$7.22\cdot10^{-3}$ (0.03)	$7.85\cdot10^{-2}$ (0.60)
CE	0.20 (1.27)	$6.34\cdot10^{-3}$ (0.16)	$1.81\cdot10^{-2}$ (0.90)	$6.06\cdot10^{-4}$ (0.34)	$1.28\cdot10^{-2}$ (2.47)*	$1.74\cdot10^{-2}$ (4.36)**	0.16 (0.62)	0.12 (0.91)
DE	$6.55\cdot10^{-2}$ (0.42)	$1.85\cdot10^{-2}$ (0.47)	$2.88\cdot10^{-5}$ (0.001)	$1.99\cdot10^{-5}$ (0.01)	$5.35\cdot10^{-4}$ (0.10)	$9.62\cdot10^{-4}$ (0.24)	$4.80\cdot10^{-2}$ (0.18)	$7.68\cdot10^{-5}$ (0.0006)

* P value between 0.05 and 0.01.
**P value between 0.01 and 0.001.
***P value lower than 0.001.

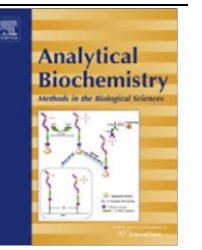
Chapter 11:

Analytical platform for verification and quantitation of target peptides in human serum: Application to cathelicidin



Analytical Biochemistry

415 (2011) 39–45



Analytical platform for verification and quantitation of target peptides in human serum: Application to cathelicidin

M. Calderón-Santiago [a,b], *J.M. Mata-Granados* [b,c], *F. Priego-Capote* [a,b*], *J.M. Quesada-Gómez* [b,c], *M. D. Luque de Castro*[a,b]

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

[c]*Sanyres I+D+I Department (Grupo Prasa), E-14012. Córdoba, Spain.*

Analytical platform for verification and quantitation of target peptides in human serum: Application to cathelicidin

M. Calderón-Santiago, J.M. Mata-Granados, F. Priego-Capote, J.M. Quesada-Gómez, M. D. Luque de Castro*

Abstract

A selective and sensitive, fully automated platform for verification and quantitative determination of target peptides in biofluids is proposed and then validated by development of a method for analysis of cathelicidin in human serum. The method is based on the on-line coupling of solid-phase extraction (SPE) and tandem mass spectrometry with direct infusion. Mass spectrometry analysis was carried out by multiple reaction monitoring using three transitions (one for quantitative analysis and two for qualitative analysis), all them confirmed by in silico fragmentation of the target peptide. Samples were prepared in the SPE workstation on a polymeric divinylbenzene resin by preconcentration, deproteinization, and cleanup, removing salts and interferences after direct injection of human serum. The analytical process required 12 min. The limits of detection and quantitation were 2.5 and 8.25 μg/L, respectively (0.20 and 0.66 pg on column). Repeatability and within-laboratory reproducibility were 2.4% and 2.7%, respectively. A dual-cartridge configuration was used to test recovery of cathelicidin in serum, resulting in 80%. Because quantitative retention in the cartridge was assessed, determination of cathelicidin was validated without using synthetic peptides labeled with stable isotopes. The hyphenated system allows full automation, thereby improving reproducibility and accuracy, as demanded by clinical analysis.

1. Introduction

Targeted mass spectrometry (MS)[1] is emerging as an analytical technique capable of selective and sensitive detection and quantitation of any protein/peptide of interest (or modification thereof) [1–4]. In selected reaction monitoring (SRM) mode, peptides (precursors) from target proteins are selectively detected and caused to fragment (products) in the mass spectrometer. The resulting product ions are used for selective quantitation of the peptide and, therefore, the protein from which it was derived. This two-stage filtering process allows chemical background to be overcome by improving the signal-to-noise ratio and permits selective determination of target peptides.

The first step in developing an SRM-based assay involves the selection of a subset of peptides as quantitative candidates for each protein. Signature peptides correspond to the subset of proteotypic peptides that are also the highest responding peptides for each protein in terms of sensitivity [5]. The major challenge for SRM is to decide the target peptides to be monitored taking into account that each protein has multiple enzymatic cleavage sites. The selection of signature peptides is currently based on detection in the preliminary MS data [1,3], identification in databases containing MS experimental data [6,7] or the use of computational approaches designed to predict proteotypic peptides [8,9]. This can be done manually, for example, by theoretic digestion of the sequences of interest and blasting of all resulting peptides (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) [10]. After selecting signature peptides, the targeted SRM–MS assay should be independently optimized for each peptide to select the appropriate precursor-to-product ion transitions [11].

SRM has proven to be a successful method for discovery and validation of novel biomarkers [1,12,13], and in comparison with other alternatives such as enzyme-linked immunosorbent assays (ELISAs), it is cost-effective, quicker to design, and suitable for multiplexed analysis [4]. Mass spectrometry in SRM mode requires a high level of ion separation but not necessarily high resolution. Therefore, the instrumentation used to measure peptide SRMs is similar to that

existing in robust platforms present in analytical laboratories. Increased throughput is also possible by direct coupling of separation (via liquid chromatography [LC]) to MS [3,14]. However, automation of sample preparation is a challenge in the analysis of target peptides.

Quantitation in SRM-based approaches is mostly carried out with stable isotopes by calculating the ratio of the signal response of the endogenous peptide to a stable isotope-labeled version of the peptide spiked as an internal standard at a known concentration (AQUA [absolute quantitation] methodology) [2]. The main limitation is the cost of isotopic-labeled peptides that need to be synthesized after optimization of the protocol. Another possibility is to evaluate potential errors of the method with an internal validation step, avoiding internal standard-based quantitation. An internal standard can be highly useful, but its selection also creates a problem unless a labeled isotopic standard is used. However, the synthesis of a isotopically labeled version of peptides is expensive, particularly if long peptides are the target of the analysis. Internal validation can be achieved by implementation of a solid-phase extraction (SPE) step that validates the analytical method by estimation of the recovery factor. For this purpose, spiked and nonspiked serum samples are used to support that the method is independent of matrix interferences after optimization of each step of the SPE protocol. This matrix independence is the key aspect for this assumption. The development of robust and automatic platforms for analysis of peptide panels or with high capability for fast individual determination of them is highly desirable.

Cathelicidin or LL-37 is a peptide formed by 37 amino acids with an α-helix secondary structure. It is one of the human antimicrobial peptides that take part in the immune response. LL-37 is synthesized as an inactive peptide called hCAP-18, which is cleaved by serine proteases in the target tissue to activate the C-terminal antimicrobial peptide (LL-37) operating in the first line of defense [15]. Cathelicidin stimulates endothelial proliferation, promoting the accumulation of leukocytes in the inflammation focus or in a wound [16]. Furthermore, this peptide protects against necrotic skin and possesses a synergic

effect with β-defensins (other type of human antimicrobial peptides) [17,18]. The levels of this peptide, estimated as messenger RNA (mRNA), can be related to different diseases such as eczema herpeticum, infective cellulites, and pulmonary tuberculosis [19–21]. This relationship confers cathelicidin a promising capability as biomarker that still needs to be defined. Preliminary results have revealed that low plasma levels of cathelicidin predict increased infectious disease mortality in patients undergoing hemodialysis [22].

Despite this relationship, few methods have been reported for determination of LL-37 in human serum or plasma. Recently, LL-37 plasma levels have been estimated using a commercial ELISA kit, an expensive approach with a protocol time of 3.5 h [23]. Cathelicidin levels have also been estimated using transcriptomic approaches by determination of cathelicidin mRNA expression [20,21]. However, it has been supposed that cathelicidin levels are exclusively linked to its mRNA expression. The aim of this research was to develop an automatic analytical platform for robust targeted analysis of peptides —a platform that has been validated by analysis of cathelicidin.

2. Material and methods

2.1. Chemicals

LC–MS-grade methanol, acetonitrile, formic acid, and acetic acid were purchased from Scharlab (Barcelona, Spain). Deionized water (18 mΩ cm) from a Millipore Milli-Q water purification system was used for preparation of all aqueous solutions. LL-37 (cathelicidin) standard (97.4% purity tested by high-performance liquid chromatography [HPLC] analysis) was provided by Bachem (Weil am Rhein, Germany).

Commercial standard cathelicidin was weighed and dissolved in water according to the manufacturer's instructions to obtain a 500 mg/L stock solution. Working solutions were prepared by dilution of the appropriate volume of stock solution in water.

2.2. Blood extraction and serum isolation

Venous blood was collected in evacuated sterile tubes for whole blood hematology determination (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) and centrifuged at 4000 rpm for 10 min to isolate the serum fraction (processing within 2 h after collection). Serum was placed in a plasticware tube and stored at or below −80 °C until analysis. 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, which were supervised by the ethical review board of Reina Sofia Hospital (Cordoba, Spain) that approved the experiments. Individuals selected for this study were told to obtain consent prior to this research.

2.3. SPE–tandem mass spectrometry configuration

Fig. 1 illustrates a scheme of the configuration device. A Midas auto-sampler furnished with a 200 μl sample loop connected to an SPE workstation Prospekt-2 system (Spark Holland, Emmen, Netherlands) was used to automate SPE. The Prospekt-2 system consisted of an automatic cartridge exchanger and a high-pressure dispenser, which enabled fully automated performance of sample preparation, controlled by SparkLink version 2.10 software. Hysphere Resin GP cartridges (8 μm, 10 x 2.0 mm, Spark Holland) were used for the SPE step. The Prospekt-2 system was on-line connected to an Agilent 1200 Series LC system (Palo Alto, CA, USA), which consists of a binary pump and a vacuum degasser. The eluate was introduced directly into an Agilent 6410 triple quadrupole (QqQ) detector furnished with an electrospray ionization (ESI) source. Agilent MassHunter Workstation was the software for data acquisition, qualitative analysis, and quantitative analysis.

2.4. Analytical protocol for determination of cathelicidin

Human serum (0.5 ml) was vortexed with 0.5 ml of 15% formic acid for 30 s into an amber injection vial, which was placed in the autosampler. The resulting solution (200 μl) was injected into the SPE–tandem mass spectrometry

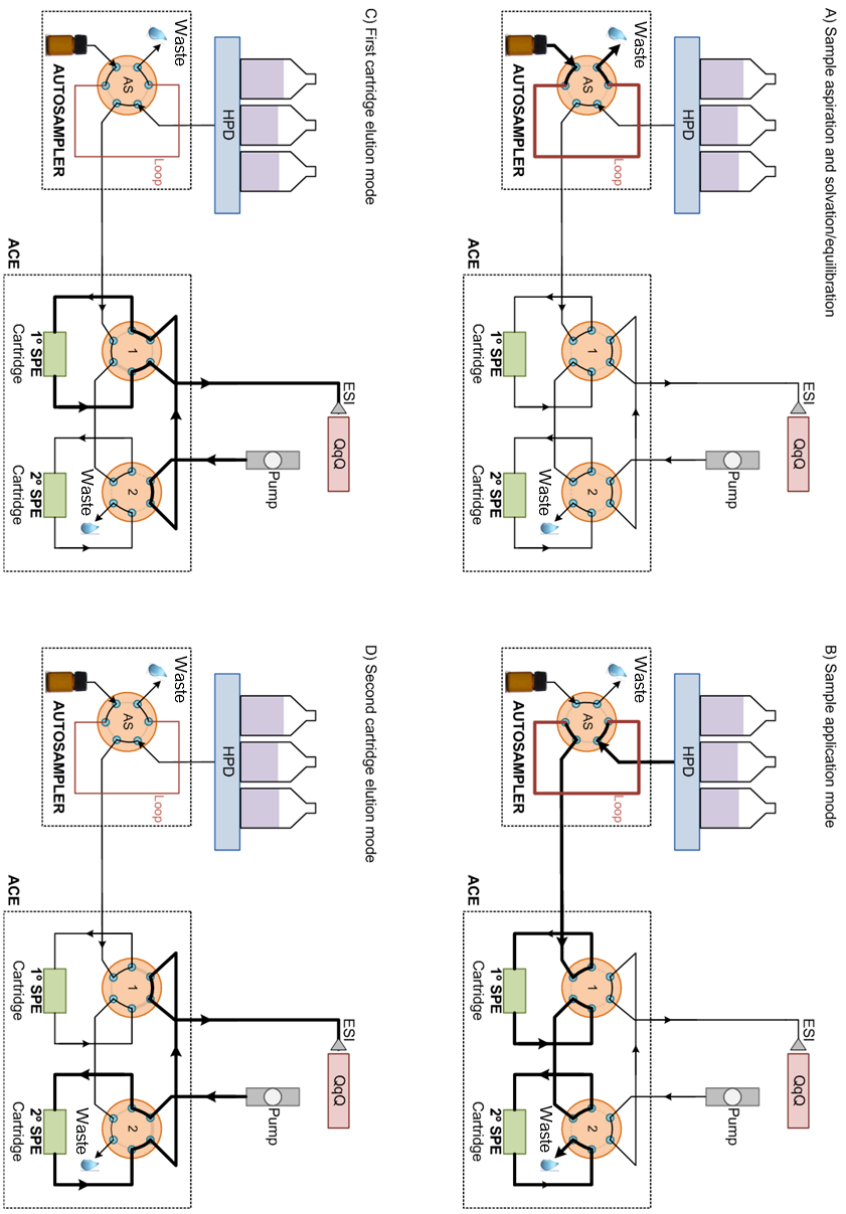


Fig. 1. Scheme of the configuration device based on a Prospekt–MS/MS coupling. ACE, automatic cartridge exchange; ESI, electrospray ionization module; HPD, high pressure

(MS/MS) system following a sequence of automatic operations. Basically, the sample preparation step starts with methanol activation of the stationary phase with subsequent equilibration before sample loading with 2 ml of aqueous formic acid. Under these conditions, LL-37 cathelicidin is retained in the cartridge, which is washed with 10% methanol/0.1% formic acid aqueous solution to remove potential interferences. Elution starts by switching the left clamp valve and pumping 40% acetonitrile/1% formic acid/2% acetic acid solution with the LC pump at 1.0 ml/min. The SPE step finishes by purging the tubes of the Prospekt-2 system. Cartridges were reused for three replicates of each sample.

MS detection was performed in positive ESI mode at unit resolution in both quadrupoles. The ESI parameters were set as follows: 5.3 kV capillary voltage, 315 °C source temperature, and 60 psi pressure nebulizer. Nitrogen gas was flowed at 10 ml/min to dry the eluent. The precursor ion for cathelicidin was 749.6 *m/z* with a charge state of $z = 6$, which was efficiently filtered by setting the voltage of the first quadrupole at 220 V. Optimal collision energy of 30 eV activated the precursor ion to generate product ions 854.0, 876.6, and 819.5 m/z selected as quantitation and qualifier ions by multiple reaction monitoring (MRM). The dwell time was fixed at 200 ms for all SRM transitions. The entire analytical process was completed in 12 min.

2.5. In silico elucidation of SRM transitions

The *m/z* value of the first quadrupole was determined by the mass (4492.58 Da) and the predominant charge state of a peptide. The mass of the fragment ions was theoretically estimated by in silico fragmentation of the precursor ion using the Protein Prospector MS product website developed by the University of California, San Francisco (http://prospector2.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct). The parameters considered for this fragmentation were quadrupole-based fragmentation, predominant generation of b and y ions [24,25] with $-H_2O$ and $-NH_3$ neutral losses, and $+H_2O$ peeling sequence, maximum charge state $z = 9$, and ±0.6 Da mass fragment tolerance.

2.6. Validation of the transitions selected for analysis

The three most intense transitions were selected for qualitative and quantitative analysis of cathelicidin peptide. These transitions were subsequently detected by search in the profile of theoretic fragment ions generated by in silico digestion of the precursor peptide.

2.7. SRM-based quantitation

The standard addition method was used to obtain the calibration model for cathelicidin and validate the analytical method. Calibration curves (see Fig. S1 in Supplementary material) were constructed by spiking human serum with known amounts of cathelicidin solutions. This calibration model was selected to correct matrix effects occurring during sample preparation (recovery, saturation effects) and analysis (suppression). Calibration solutions were prepared from cathelicidin stock solutions (25 and 500 mg/L) at 25, 50, 75, 100, 200, and 500 μg/L. Three of them were injected in triplicate. In all cases, the samples were human serum spiked with the target analyte; therefore, another calibration level corresponded to 0 μg/L standard addition.

3. Results and discussion

3.1. Development of the MS protocol

Preliminary experiments can be exploited to derive information about the predominant precursor charge state and the MS/MS fragmentation pattern of a target peptide. However, it is important to be aware that the ionization conditions can affect distribution of the charge state. Ionization optimization was initially focused on the selection of the preferred ESI mode for MS analysis of cathelicidin. This was performed by direct injection of standard solutions using both positive and negative ESI modes. The highest sensitivity was achieved with positive ESI mode that generated different charge states and, therefore, different *m/z* signals. The influence of the ionization agent was studied by the addition of different concentrations of trifluoroacetic acid, acetic acid, and formic acid. The

best results were obtained by a combination of 1–2% formic acid/acetic acid in a 30% methanol aqueous solution. The electrospray variables were set by a multivariate response surface design consisting of 16 experiments and two central points (experimental runs in which the value of each factor is the median of the values used in the factorial portion) by monitoring the ion indicated previously. The ranges studied for the temperature, pressure, and capillary voltage were 185–315 °C, 10–60 psi, and 2700–5300 V, respectively. The effect of these three variables in the signal intensity corresponding to the target ion was positive, so all of the variables were established at their maximum values.

The *m/z* value of the first quadrupole was determined by the mass (4492.58 Da) and the predominant charge state of a peptide, which was $z = 6$, as shown in the mass spectrum illustrated in Fig. 2, obtained in full scan mode by direct injection of an LL-37 standard. In the mass range under study (0–2000 Da), the charge distribution ranged from $z = 3$ to 6. Therefore, the ion 749.6 *m/z* with charge state $z = 6$ was selected for quantitative purposes.

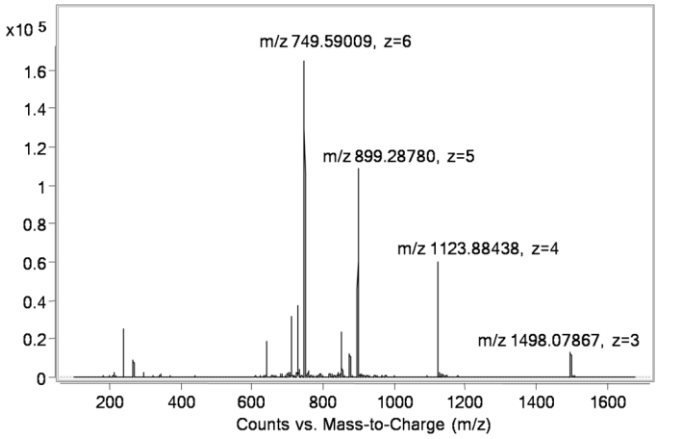


Fig.2. Mass scan in the range of 0–2000 Da by direct infusion of a cathelicidin standard showing the signals corresponding to the most stable charge states.

The mass of the product ions was theoretically estimated by in silico fragmentation of the precursor ion using the Protein Prospector MS product website. Table S1 (see Supplementary material) shows the sequence of *b* and *y* ions generated with fragmentation in the hexapole collision cell, including $-H_2O$

and $-NH_3$ neutral losses and $+H_2O$ peeling sequence. A maximum charge state $z = 9$ was set with ±0.6 Da mass fragment tolerance. It is worth emphasizing that the transitions selected for qualitative and quantitative analysis of LL-37 should be selective enough to develop a highly selective and sensitive assay. The distribution of relative fragment ion intensities is dependent on the type of instrument used and the operating parameters. This is particularly relevant when ion-trap-derived data are used to select transitions for a QqQ instrument. Owing to the different mode of collision-induced activation in ion traps compared with quadrupole collision cells, higher *b*-type ions and doubly charged fragments are usually less prominent or absent in the QqQ instrument mass spectra. Product ions with greater *m/z* than that of the precursor ion are preferred because they fit in with a greater proportion of the original peptide in the spectrum; therefore, the selectivity is significantly increased by these transitions [26,27].

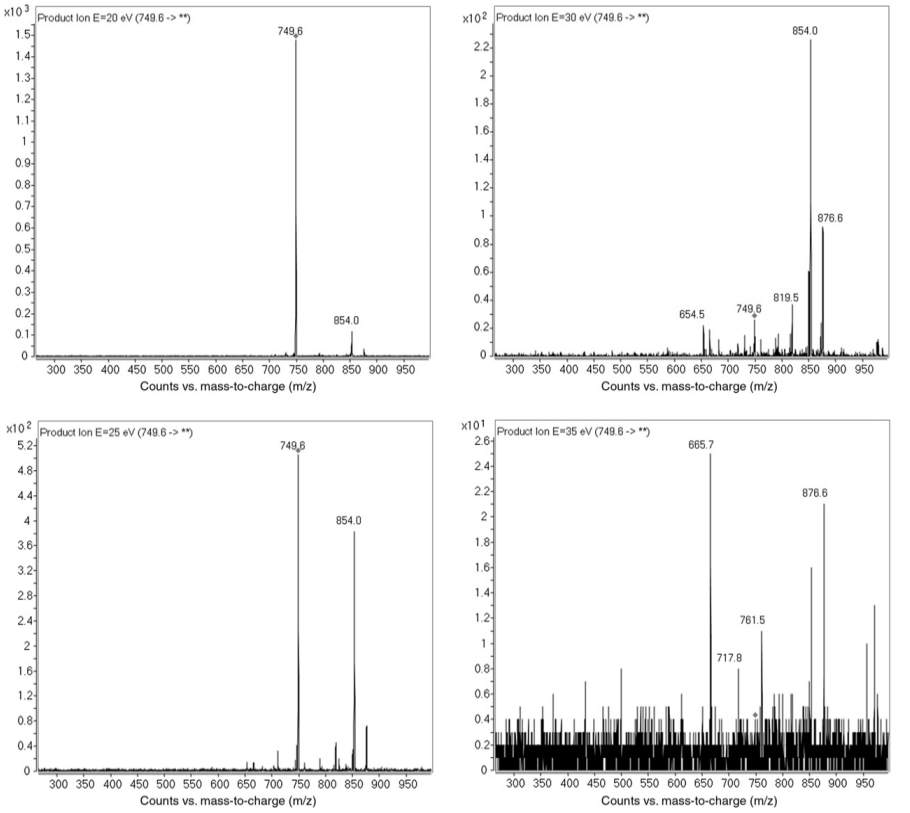


Fig. 3. MS/MS spectra generated by fragmentation of cathelicidin precursor ion 749.6 m/z with collision energies from 15 to 45 eV.

MS/MS behavior of cathelicidin was studied using different collision energy values from 15 to 45 eV. Fig. 3 shows the MS/MS spectra where the fragmentation dependence on the collision energy applied in the cell can be observed. Fragmentation of the precursor ion is not significant up to 25 eV by generation of different product ions at 819.5, 854.0, and 876.6 *m/z* that theoretically fit for ions y_{33}^{5+}, y_{35}^{5+}, and y_{36}^{5+}, respectively, according to Supplementary Table S1. Collision energy of 30 eV was the optimal value, as deduced from Fig. 3 by virtue of the efficient fragmentation confirmed by the intensity ratio of the signals corresponding to precursor and product ions. Higher values of collision energy led to an exhaustive fragmentation of the precursor ion without a significant increase of the product ion signal. The product ion at 854.0 *m/z* was selected for quantitation, and the transitions to 819.5 and 876.6 *m/z* were selected as qualifiers for verification. Fig. 4A shows the result of the analysis of a 50 ng/ml cathelicidin standard by direct infusion with MS/MS detection in SRM mode. In simple protein mixtures, a single transition may be sufficient to monitor a particular protein of interest, but in complex samples such as serum and other biofluids, multiple transitions per peptide are generally required because of the interferences caused by background and peptides coming from high-abundant proteins.

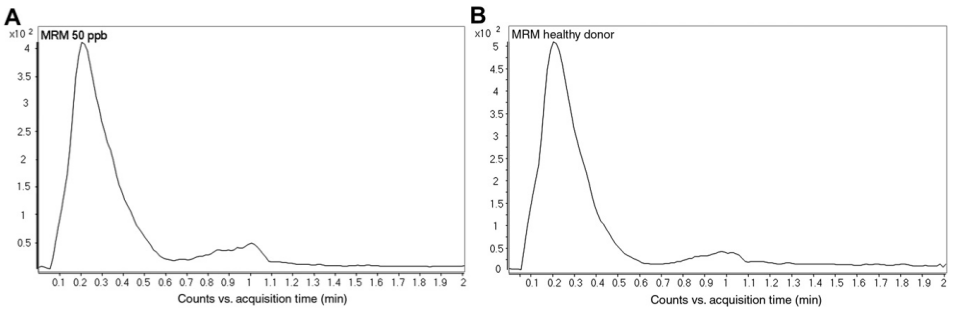


Fig. 4. SRM analysis provided by a 50-ng/L cathelicidin standard (A) and a serum sample from a healthy donor (B).

3.2. Development of the SPE protocol

After optimization of the detection step, automatic SPE was optimized. All variables were studied by a univariate mode due to the inclusion of noncontinuous variables such as composition of loading, washing, or elution solutions. The response variables for optimization of the SPE step were both the area and height of the peak signal to maximize sensitivity and minimize dispersion. The SPE step was validated by using a dual configuration with two cartridges to estimate the recovery factor (calculated as the analyte fraction retained in the first cartridge), which was the response variable of this optimization study. Serum spiked with LL-37 was used to select the most efficient SPE sorbent, and the rest of variables, the range within which they were studied, and the optimal values are shown in Table 1. Concerning the SPE cartridge, eight types of SPE sorbent with different retention properties were tested: Hysphere CN (silica-based cyanopropyl phase), Hysphere C2 (silica-based ethyl phase), Hysphere C8 (silica-based octyl phase), Hysphere C8 (EC) (end-capped silica-based octyl phase), Hysphere C18 (silica-based octadecyl phase), Hysphere C18 HD (high-density silica-based octadecyl phase), Hysphere Resin GP (polymeric polydivinylbenzene phase), and Hysphere Resin SH (strong hydrophobic-modified polystyrene-divinylbenzene). The best results, in terms of pre-concentration and cleanup, were obtained with the Hysphere Resin GP cartridge, which showed higher cathelicidin retention, the highest recovery, and a homogeneous elution. Another aspect of the optimization study was focused on the composition and pH of the different solutions used in the SPE protocol for loading the sample and washing the cartridge for efficient removal of interferents and salts and LL-37 elution. The composition of the solutions was studied by the addition of organic solvents in a wide concentration range to favor cathelicidin retention, cleanup, and elution. The three operations were also favored by acid pHs, as Table 1 shows. In addition, other variables studied were volume and flow rate of the different solutions as well as the elution time, which was controlled by switching the valve clamp.

Table 1. Optimization of main variables involved in SPE step.

Variable		Tested range		Optimal conditions	Conditions employed
SPE sorbent		Hysphere CN, C2, C8, C8-EC, C18, C18-HD, Resin GP and Resin SH		Hysphere Resin GP, Resin SH and C8	Hysphere Resin GP
Loading solvent	Composition	Organic proportion in the aqueous phase	0–50% methanol and acetonitrile	0% organic phase	15% formic acid
		Acidification	0–15% formic acid and acetic acid	15% formic acid	
	Volume	0.5–4 ml		2 ml	2 ml
	Flow Rate	0.5–3 ml/min		0.5 ml	0.5 ml
Washing solvent	Composition	Organic proportion in the aqueous phase	0–50% methanol and acetonitrile	10% methanol	10% methanol with 0.1% formic acid
		Acidification	0.1–10% of formic acid	0.1% formic acid	
	Volume	0.5–4 ml		2 ml	2 ml
	Flow rate	1–3 ml/min		1 ml/min	1 ml/min
Elution solvent	Composition	Organic proportion in the aqueous phase	Acetonitrile, methanol, tetrahydrofurane, and isopropanol in different proportions	40% acetonitrile	40% acetonitrile
		pH for optimal cathelicidin ionization	0.5–2% formic acid and acetic acid	1% formic acid and 2% acetic acid	1% formic acid and 2% acetic acid
	Elution time	0.5–2 min		1 min	1 min
	Flow rate	0.4–1 ml/min		1 ml/min	1 ml/min

A key factor was the dilution of the serum samples to improve the loading step, minimize proteins interactions, and maximize cathelicidin retention in the cartridge, thereby increasing cathelicidin recovery. Water, different concentrations of formic acid, and 50 mM phosphate buffer in the pH range 3.0 to 7.5 were used as solutions for dilution from 50% to 90%. The best results were obtained with 15% aqueous formic acid diluting up to 50% (7.5% final formic acid concentration in the sample). Fig. 4B illustrates the result of the analysis of a human serum sample with SPE as sample preparation; it was highly efficient, as demonstrated by comparison with the analysis of an LL-37 standard. The automated protocol developed with the SPE workstation is schemed in Table 2, which also includes the time required for the different steps.

Table 2. Scheme of automated SPE protocol for preconcentration and cleanup of cathelicidin in human serum prior to on-line elution to mass spectrometer.

Step		Solvent	Volume (ml)	Flow (ml/min)	Duration
1	New cartridge and Star autosampler				
2	Solvation	Methanol	1	3	0 min 26 s
3	Equilibration	15% formic acid	2	3	0 min 52 s
4	Equilibration	15% formic acid	1	1	1 min 06 s
5	Sample application	15% formic acid	2	0.5	4 min 12 s
6	Wash cartridge	10% methanol and 0.1% formic acid	2	1	2 min 12 s
7	Elution	40% acetonitrile, 1% formic acid, and 2% acetic acid	1	1	1 min 00 s
8	Wash cartridge	Methanol	2	3	0 min 52 s
9	Wash cartridge	Water	4	3	1 min 44 s
Total time					**12 min 36 s**

3.3. Validation of the method

Analytical characterization of the method was carried out with a serum pool from healthy individuals. The lowest limit of detection (LLOD) and lowest limit of quantitation (LLOQ) were determined by injecting dilution series of cathelicidin to obtain the concentration with a signal 3 and 10 times the noise (average noise value obtained for blank injections in the region of the spectrum of interest), respectively, resulting in LLOD and LLOQ values of 2.5 and 8.25 μg/L, respectively (0.20 and 0.66 pg/injection). Calibration curves were established by applying the method of standard additions using cathelicidin stock solutions. Different levels of the target peptide (25–500 μg/L) were added to pool aliquots to plot signal responses versus the spiked concentration. The correlation coefficient was 0.998, and the linear dynamic range was from 8.25 to 500 μg/L.

The accuracy of the method and potential matrix effects were assessed by analysis of nonspiked and spiked serum samples at a high concentration level (100 μg/L) to evaluate the retention capability of the cartridge. The recovery of cathelicidin was calculated with the two-cartridge configuration of the Prospekt-2 system by analysis of five replicates [28]. Two Hysphere Resin GP cartridges were located in serial, so after sample injection the amount of cathelicidin not retained

in the first cartridge was retained in the second. The eluates from both cartridges were sequentially injected into the QqQ analyzer, estimating the concentration retained in each cartridge. The recovery factor in this system was calculated as amount retained in cartridge 1/(amount retained in cartridge 1 + amount retained in cartridge 2), resulting in 80% recovery. This study was carried out at a high concentration of cathelicidin as compared with the normal range. Nevertheless, this recovery was confirmed by repetition of the test with samples from the individual cohort selected for validation of the method.

Within-laboratory reproducibility and repeatability were evaluated in a single experimental setup with duplicates by experiments carried out with a serum pool for 1 week [29]. The repeatability, expressed as relative standard deviation (RSD), was 2.4%, and the within-laboratory reproducibility, also expressed as RSD, was 2.7%.

3.4. Application of the method to human serum samples

Serum samples isolated from intensive care patients (n = 17) and healthy blood donors as a control group (n = 23) were collected (Regional Blood Donors Center, Córdoba, Spain) to compare LL-37 levels and establish normal concentrations in human serum. The mean age of the population was 41 ± 16 years formed by 26 males and 14 females individuals. The samples were injected in triplicate, being an RSD lower than that estimated in the repeatability test. As can be seen in Fig. 5, the group formed by intensive care patients possessed a higher variability in LL-37 profile as compared with the group of control individuals, which can be represented by a normal distribution. In addition, the average LL-37 level was higher in intensive care patients (48.4 ± 29.8 ng/ml) than in the control group (34.9 ± 18.7 ng/ml). However, no statistical differences were found with a running t test (P = 0.22) taking into account standard deviation values illustrated in Fig. 5. These preliminary results enable us to conclude only that cathelicidin levels were within a controlled range in healthy individuals, whereas significant variability was found in intensive care patients. In this research, a reduced variable cohort was analyzed to study cathelicidin levels. A subsequent

step would be to study a more representative cohort classified according to pathological states to discriminate between them. This could be the initial step to evaluate the potential of cathelicidin as a clinical biomarker for different pathologies.

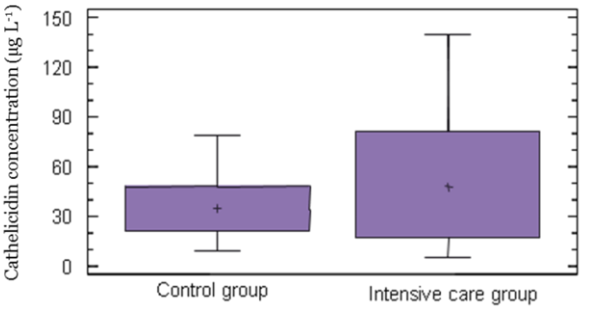


Fig. 5. Cathelicidin levels found in intensive care patients versus a control group of healthy individuals.

4. Conclusions

An automated platform has been proposed for verification and quantitative analysis of target peptides in biofluids. The method is based on direct injection of the biofluid into an automated SPE device where sample preparation is efficiently carried out by desalting and deproteinization. This is one of the main differences compared with other devices existing in the literature [30], especially those designed for biochemical assays developed with standard solutions and, therefore, not useful for clinical analysis of biofluids. In this way, the resulting eluate is eluted directly into the mass spectrometer with highly efficient ionization without affecting electrospray performance. The method was applied to the analysis of cathelicidin peptide in human serum. The analyte was detected by MS/MS in SRM mode using three different transitions for verification and quantitative determination. The use of an internal standard was avoided by SPE validation using tests with dual cartridge configurations to check quantitative retention of cathelicidin. In this way, sample spiking with synthetic peptides labeled with stable isotopes, which notably increases the cost of the

analysis, is not required. The overall method was properly validated by application to human serum samples in order to set the average concentration of cathelicidin peptide in healthy individuals. This is another benefit versus single-cartridge configurations that omit internal validation tests [31].

The approach presented here could be used as a template for quantitative analysis of other peptides such as clinical biomarkers and target proteins after enzymatic hydrolysis to generate representative peptides. Thus, with minor or no changes, it can be applied to any other peptide with a considerably reduced cost per analysis, complete automation without implementation of robotized work-stations, and an extra level of selectivity to suppress cross-reactivity.

Acknowledgements

The Spanish Ministerio de Ciencia e Innovación (MICINN) is thanked for financial support through Project CTQ2009-07430. F. Priego-Capote is also grateful to the MICINN for a Ramón y Cajal Contract (RYC-2009-03921). M. Calderón-Santiago also thanks the "Initiation to Research Program," University of Córdoba, for a fellowship. The Regional Blood Donors Center in Córdoba, Spain, is thanked for providing serum samples.

References

[1] L. Anderson, C.L. Hunter, Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins, Mol. Cell. Proteomics 5 (2006) 573–588.

[2] S.A. Gerber, J. Rush, O. Stemman, M.W. Kirschner, S.P. Gygi, Absolute protein expresión profiling estimates the relative contribution of transcriptional and translational regulation, Proc. Natl. Acad. Sci. USA 100 (2003) 6940–6945.

[3] H. Keshishian, T. Addona, M. Burgess, E. Kuhn, S.A. Carr, Isotope dilution-mass spectrometric quantification of specific proteins: model application with apolipoproteinA-I, Mol. Cell. Proteomics 6 (2007) 2212–2229.

[4] J. Stahl-Zeng, V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold, High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites, Mol. Cell. Proteomics 6 (2007)1809–1817.

[5] B. Kuster, M. Schirle, P. Mallick, R. Aebersold, Scoring proteomes with proteotypic peptide probes, Nat. Rev. Mol. Cell. Biol. 6 (2005) 577–583.

[6] R. Craig, J.P. Cortens, R.C. Beavis, Use of multiple peptide identification programs to increase peptide and protein identifications in human whole saliva, J. Proteome Res. 3 (2004) 1234–1242.

[7] E.W. Deutsch, H. Lam, R. Aebersold, Peptide Atlas: a resource for target selection for emerging targeted proteomics workflows, EBMO Reports 9 (2008) 429–434.

[8] P.T. Van, A.K. Schmid, N.K. King, A. Kaur, M. Pan, K. Whitehead, T. Koide, M.T. Facciotti, Y.A. Goo, E.W. Deutsch, D.J. Reiss, P. Mallicj, N.S. Baliga, HalobacteriumSalinarum NRC-1. Peptide Atlas: Toward strategies for targeted proteomics and improved prometome coverage, J. Proteome Red. 7(9) (2008) 3755–3764.

[9] W.S. Sanders, S.M. Bridges, F.M. McCarthy, B. Nanduri, S.C. Burgess, Prediction of peptides observable by mass spectrometry applied at the experimental set level, BMC Bioinf. 8 (2007) S7–S23.

[10] M. Kohl, G. Redlich, M. Eisenacher, A. Schnabel, H.E. Meyer, K. Marcus, C. Stephan, Automated calculation of unique peptide sequences for unambiguous identification of highly homologous proteins by mass spectrometry, J. Proteomics Bioinf. 1(2008) 6–10.

[11] J.D. Jaffe, H. Keshishian, B. Chang, T.A. Addona, M.A. Gillette, S.A. Carr, Accurate inclusion mass screening: a bridge from unbiased discovery to

targeted assay development for biomarker verification, Mol. Cell. Proteomics 7 (2008) 1952–1962.

[12] N. Rifai, M.A. Gillette, S.A. Carr, Protein biomarker discovery and validation: the long and uncertain path to clinical utility, Nat. Biotechnol. 4 (2006) 971–983.

[13] E. Kuhn, J. Wu, J. Karl, H. Liao, W. Zolg, B. Guild, Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards, Proteomics 4 (2004) 1175–1186.

[14] R.G. Kay, B. Gregory, P.B. Grace, S. Pleasance, The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum, Rapid Commun. Mass Spectrom. 21 (2007) 2585–2593.

[15] O.E. Sørenser, P. Follin, A.H. Johnsen, J. Calafat, G.S. Tjabringa, P.S. Hiemstra, N. Borregaard, Human cathelicidin, hCAP-18, is processed to the antimicrobial peptide LL-37 by extracellular cleavage with proteinase 3, Blood 97 (2001) 3951–3959.

[16] L.E. Castrillón Rivera, A. Palma Ramos, D. Padilla Desgarennes, La función inmunológica de la piel, Dermatol. Rev. Mex. 51 (2007) 57–67.

[17] T. Ganz, Defensins: antimicrobial peptides of innate immunity, Nat. Rev. Immunol. 3 (2003) 711–720.

[18] T. Ganz, The role of antimicrobial peptides in innate immunity, Integr. Comp. Biol. 43 (2003) 300–304.

[19] M.D. Howell, A. Wollenberg, R.L. Gallo, M. Flaig, J.E. Streib, C. Wong, T. Pavicic, M. Boguniewicz, D.Y.M. Leung DYM, Cathelicidin deficiency predisposes to eczema herpeticum, J. Allergy Clin. Immunol. 117 (2006) 836–841.

[20] M.E. Stryjewski, R.P. Hall, V.H. Chu, Z.A. Kanafani, W.D. O'Riordan, M.S. Weinstock, R.S. Stienecker, R. Streilein, R.A. Dorschner, V.G. Fowler Jr, G.R. Corey, R.L. Gallo, Expression of antimicrobial peptides in the normal and involved skin patients with infective cellulitis, J. Infect. Dis. 196 (2007) 1425–1430.

[21] P. Selvaraj, S. Prabhu Anand, M. Harishankar, K. Alagarasu, Plasma 1,25 dihydroxy vitamin D3 level and expression of vitamin D receptor and cathelicidin in pulmonary tuberculosis, J. Clin. Immunol. 29 (2009) 470–478.

[22] A.F. Gombart, I. Bhan, N. Borregaard, H. Tamez, C.A. Camargo, H.P. Koeffler, R. Thadhani, Clin. Infectious Diseases 48 (2009) 4818–4824.

[23] L. Jeng, A.V. Yamshchikov, S.E. Judd, H.M. Blumberg, G.S. Martin, T.R. Ziegler, V.L. Tangpricha, Alterations in vitamin D status and anti-microbial peptide levels in patients in the intensive care unit with sepsis, J. Transl. Med. 7 (2009) 28–36.

[24] P. Roepstorff, J. Fohlman, Letter to the editors, Bio. Mass Spec. 11 (1984) 601.

[25] R. S. Johnson, S. A. Martin, K. Biemann, J. T. Stults, J. T. Watson, Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine, Anal. Chem. 59 (21) (1987) 2621–2625.

[26] X. Yang, I.M. Lazar, MRM screening/biomarker discovery with linear ion trap MS: a library of human cancer-specific peptides, BMC Cancer 9 (2009) 96–106.

[27] V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantatitive proteomics: a tutorial, Mol. Syst. Biol. 4 (2008)1–14.

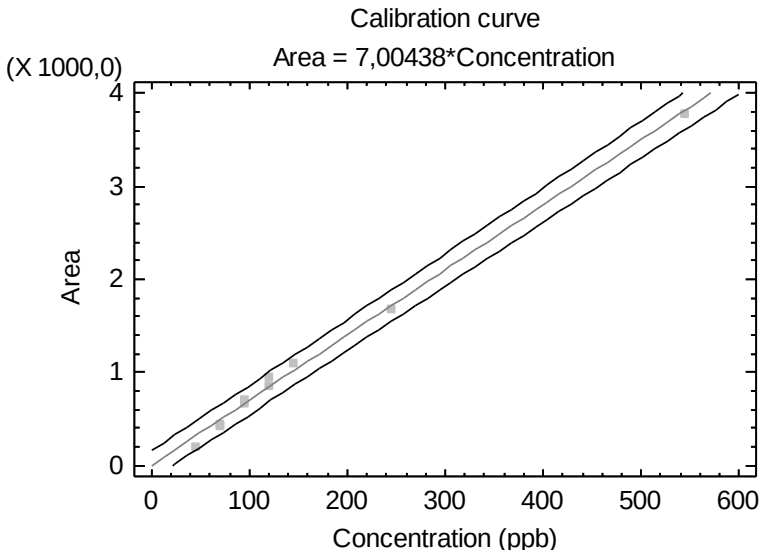
[28] J.A. Bert Ooms, G.J. Mark Van Gils, A.R. Duinkerken, O. Halmingh, Am. Lab. 32 (2000) 52–57.

[29] D. L. Massart, B. G. M. Vanderginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, J. Smeyers-verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Amsterdam, 1997.

[30] K. B. Lim, C. C. Özbal, D. B. Kassel, Development of a high-throughput online solid-phase extraction/tandem mass spectrometry method for cytochrome P450 inhibition screening, J. Biomol. Screen 15(4) (2010) 447–452.

[31] M. W. J. van Hout, C. M. Hofland, H. A. G. Niederländer, G. J. de Jong, On-line coupling of solid-phase extraction with mass spectrometry for the analysis of biological samples. II. Determination of clenbuterol in urine using multiple-stage mass spectrometry in an ion-trap mass spectrometer, Rapid Comm. Mass Spec. 14(22) (2000) 2103–2111.

Supplementary material



Supplementary Fig. 1. Calibration curve obtained for cathelicidin using standard addition method.

Supplementary Table 1. In silico fragmentation of cathelicidin (LL-37).

L(1.0)LGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES

User AA Formula 1: **C2 H3 N1 O1**

MH^{+1}(av)	MH^{-1}(mono)	MH^{-2}(av)	MH^{-2}(mono)	MH^{-3}(av)	MH^{-3}(mono)	MH^{-4}(av)	MH^{-4}(mono)	MH^{-5}(av)	MH^{-5}(mono)	MH^{-6}(av)	MH^{-6}(mono)	MH^{-7}(av)	MH^{-7}(mono)	MH^{-8}(av)	MH^{-8}(mono)	MH^{-9}(av)	MH^{-9}(mono)
4495.2900	4492.5827	2248.1487	2246.7950	1499.1016	1498.1991	1124.5781	1123.9011	899.8639	899.3224	750.0545	749.6032	643.0478	642.6609	562.7928	562.4542	500.3722	500.0712

[–] **Main Sequence Ions**

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
		L (1.0)	L	G	D	F	F	R	K	S	K	E	K	I	G	K	E	F	K	R	I	V	Q	R	I	K	D	
		37	36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11
C-terminal																												
y	---	4378.4986	4265.4146	4208.3931	4093.3662	3946.2978	3799.2293	3643.1282	3515.0333	3428.0012	3299.9063	3170.8637	3042.7687	2929.6847	2872.6632	2744.5682	2615.5256	2468.4572	2340.3623	2184.2611	2071.1771	1972.1087	1844.0501	1687.9490	1574.8649	1446.7700	1331
y^{-2}	---	2189.7530	2133.2109	2104.7002	2047.1867	1973.6525	1900.1183	1822.0678	1758.0203	1714.5043	1650.4568	1585.9355	1521.8880	1465.3460	1436.8352	1372.7878	1308.2665	1234.7322	1170.6848	1092.6342	1036.6922	986.5580	922.5287	844.4781	787.9361	723.8886	666
y^{-3}	---	1460.1711	1422.4764	1403.4692	1365.1269	1316.1041	1267.0813	1215.0476	1172.3493	1143.3386	1100.6403	1057.6261	1014.9278	977.2331	958.2259	915.5276	872.5134	823.4906	780.7923	728.7586	691.0639	658.0411	615.3549	563.3212	525.6265	482.9282	444
y^{-4}	---	1095.3801	1067.1091	1052.8537	1024.0970	987.3299	950.5628	911.5375	879.5138	857.7558	825.7320	793.4714	761.4476	733.1766	718.9213	686.8975	654.6369	617.8698	585.8460	546.8207	518.5497	493.7826	461.7680	422.7427	394.4717	---	---
y^{-5}	---	876.5055	853.8887	842.4844	819.4791	790.0654	760.6517	729.4315	703.8125	686.4061	660.7871	634.9786	609.3596	586.7428	575.3385	549.7195	523.9109	494.4973	468.8783	437.6581	415.0412	395.2276	369.6158	---	---	---	---
y^{-6}	---	730.5892	711.7418	702.2382	683.0671	658.5557	634.0443	608.0274	586.6783	572.1729	550.8238	529.3167	507.9675	489.1202	479.6166	458.2674	436.7603	412.2489	390.8998	---	---	---	---	---	---	---	---
y^{-7}	---	626.3632	610.2083	602.0624	585.6300	564.6202	543.6104	521.3103	503.0110	490.5778	472.2786	453.8439	435.5446	419.3898	411.2438	392.9446	374.5099	353.5001	---	---	---	---	---	---	---	---	---
y^{-8}	---	548.1937	534.0582	526.9305	512.5521	494.1686	475.7850	456.2724	440.2605	429.3815	413.3697	397.2393	381.2275	367.0919	359.9643	---	---	---	---	---	---	---	---	---	---	---	---	---
y^{-9}	---	487.3952	474.8303	468.4946	455.7138	439.3729	423.0320	405.6874	391.4546	381.7844	367.5516	353.2135	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
y-NH_3	---	4361.4721	4248.3880	4191.3666	4076.3396	3929.2712	3782.2028	3626.1017	3498.0067	3410.9747	3282.8797	3153.8371	3025.7422	2912.6581	2855.6366	2727.5417	2598.4991	2451.4307	2323.3357	2167.2346	2054.1505	1955.0821	1827.0235	1670.9224	1557.8384	1429.7434	1314
y-NH_3^{-2}	---	2181.2397	2124.6977	2096.1869	2038.6734	1965.1392	1891.6050	1813.5545	1749.5070	1705.9910	1641.9435	1577.4222	1513.3747	1456.8327	1428.3220	1364.2745	1299.7532	1226.2190	1162.1715	1084.1209	1027.5789	978.0447	914.0154	835.9649	779.4228	715.3753	657
y-NH_3^{-3}	---	1454.4935	1416.8009	1397.7937	1359.4514	1310.4286	1261.4058	1209.3721	1166.6738	1137.6631	1094.9648	1051.9506	1009.2522	971.5576	952.5504	909.8521	866.8379	817.8151	775.1168	723.0831	685.3884	652.3656	609.6794	557.6457	519.9510	477.2527	438
y-NH_3^{-4}	---	1091.1235	1062.8525	1048.5971	1019.8494	983.0733	946.3062	907.2809	875.2571	853.4991	821.4754	789.2147	757.1910	728.9200	714.6646	682.6409	650.3802	613.6131	581.5894	542.5641	514.2931	489.5260	457.5113	418.4861	390.2150	---	---
y-NH_3^{-5}	---	873.1002	850.4834	839.0791	816.0737	786.6601	757.2464	726.0262	700.4072	683.0008	657.3818	631.5732	605.9543	583.3374	571.9331	546.3142	520.5056	491.0920	465.4730	434.2527	411.6359	391.8222	366.2105	---	---	---	---
y-NH_3^{-6}	---	727.7514	708.9041	699.4905	680.2993	655.7179	631.2065	605.1897	583.8405	569.3352	547.9860	526.4789	505.1298	486.2824	476.7788	455.4297	433.9226	409.4112	388.0620	---	---	---	---	---	---	---	---
y-NH_3^{-7}	---	623.9308	607.7760	599.6300	583.1976	562.1878	541.1781	518.8779	500.3786	488.1455	469.8462	451.4115	433.1123	416.9574	408.8115	390.5122	372.0775	351.0678	---	---	---	---	---	---	---	---	---
y-NH_3^{-8}	---	546.0654	531.9299	524.8022	510.4238	492.0403	473.6567	454.1441	438.1322	427.2532	411.2413	395.1110	379.0991	364.9636	357.8359	---	---	---	---	---	---	---	---	---	---	---	---	---
y-NH_3^{-9}	---	485.5034	472.9385	466.6028	453.8220	437.4810	421.1401	403.7955	389.5628	379.8925	365.6598	351.3217	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

DISCUSIÓN DE LOS RESULTADOS

La normativa actual de la Universidad de Córdoba referente a la presentación de la Memoria de la Tesis Doctoral, en la modalidad en la que se incluyen los artículos (publicados o próximos a su publicación) como tales, establece que ha de incluirse una sección de discusión conjunta de los resultados.

La investigación que constituye la parte principal de la Tesis tiene como denominador común la metabolómica en el área en dos de las vertientes de esta disciplina ómica: El diagnóstico, pronóstico y tratamiento de enfermedades, y la nutrición. Ambas constituyen dos de los pilares básicos que soportan el concepto de medicina personalizada. La investigación que se presenta en esta Memoria se ha dividido en tres secciones en función, tanto del objetivo perseguido, como de la estrategia metabolómica utilizada en cada caso. Con este criterio, la Sección I recoge el desarrollo de diferentes innovaciones metodológicas orientadas a mejorar dos parámetros analíticos básicos, tales como sensibilidad y selectividad, pero también a ampliar el número de metabolitos detectados y facilitar la posibilidad de su identificación. Por su parte, las Secciones II y III se dedican a las dos estrategias básicas en metabolomica: Análisis global ("*untargeted analysis*") y análisis orientado ("*targeted analysis*"), respectivamente. En concreto, la Sección II recoge la investigación realizada sobre perfiles metabólicos ("*profiling analysis*"), que es el análisis global más utilizado en metabolómica clínica y cuyo objetivo es abarcar la mayor parte posible del metaboloma de un biofluido o un tejido dado. Por el contrario, la Sección III engloba la investigación sobre determinación de grupos específicos o familias de compuestos de interés clínico por su implicación en rutas biológicas cruciales que pueden alterarse o desregularse como consecuencia de ciertos estados patológicos. Los metabolitos estudiados y que constituyen esta sección fueron aminoácidos esenciales, compuestos implicados en el ciclo de los ácidos tricarboxílicos, fosfolípidos y un péptido antimicrobiano: La catelicidina. Los resultados obtenidos en cada una de

las tres secciones se discuten a continuación, no sin antes comentar algunos aspectos instrumentales.

El uso de la espectrometría de masas como técnica para la detección es común a las tres secciones. Las propiedades analíticas de esta técnica, en términos de sensibilidad, selectividad, exactitud, precisión y resolución, la convierten en la herramienta más útil para análisis metabolómico, tanto orientado como global, en muestras clínicas. Las dos estrategias típicas en metabolómica se benefician de esta técnica en sus diferentes modos, que son función del objetivo que se persiga: Análisis cualitativo o cuantitativo. Entre los diferentes tipos de espectrómetros de masas el más utilizado para análisis global es el QTOF gracias a su alta resolución (exactitud en la relación *m/z*), velocidad de barrido y sensibilidad a un precio competitivo. Por estas razones constituye el detector más adecuado para análisis cualitativo y también para cuantificación relativa. Por otra parte, el QqQ es tremendamente operativo en análisis orientado, por su sensibilidad y selectividad para análisis confirmatorio. Por tanto, es éste el detector de masas óptimo con fines cuantitativos. Teniendo en cuenta la complejidad de las muestras clínicas, los espectrómetros de masas son muy adecuados en metabolómica clínica, especialmente cuando se acoplan a sistemas de separación cromatográficos. La naturaleza de la técnica de separación (principalmente un cromatógrafo de gases o de líquidos) depende de las propiedades químicas de los metabolitos en estudio.

Sección I. Desarrollo metodológico e innovación en análisis metabolómico

A pesar de que los avances tecnológicos —que incluyen muy especialmente la espectrometría de masas de alta resolución— permiten la detección de metabolitos de forma fiable, existen aún algunas debilidades de la técnica en análisis metabolómico que tienen que superarse, especialmente las relacionadas con la obtención de perfiles metabólicos o análisis global.

Uno de los problemas en la obtención de perfiles metabólicos es la dificultad de abarcar la detección de todos los metabolitos mediante una única plataforma analítica, ya que el metaboloma está compuesto por una amplia variedad de metabolitos con muy diferentes estructuras químicas y presentes en un rango muy amplio de concentraciones. Otro aspecto clave es la complejidad de la mayor parte de los biofluidos generalmente utilizados en análisis clínico, tales como suero/plasma, orina o saliva; lo que dificulta aún más las aplicaciones clínicas de la metabolómica. En el desarrollo de la investigación que constituye esta Tesis Doctoral se plantearon tres retos para contribuir a la resolución de estos problemas, cuyos resultados se recogen en esta sección. Uno de ellos se refiere a la preparación de la muestra de un biofluido poco estudiado como es el sudor, y que presenta como principales ventajas su muestreo no invasivo y una composición simple si se compara con otros biofluidos tales como sangre u orina. La proposición de biofluidos alternativos en análisis clínico es de gran interés si se tiene en cuenta que los convencionales son con frecuencia poco o nada selectivos para reflejar cambios asociados a alteraciones fisiológicas. El segundo reto se orientó al desarrollo de una estrategia de preparación de la muestra con la que aumentar la detección de metabolitos en el análisis de muestras clínicas mediante LC–MS/MS. La plataforma instrumental para el desarrollo de la estrategia fue un sistema automático para SPE acoplado en línea al LC–QTOF. El dispositivo para SPE incluyó dos sorbentes con propiedades complementarias para la retención. El tercer reto fue también una innovación metodológica basada en la aplicación en metabolómica del fraccionamiento en fase gaseosa (GPF) con el que se pretendió mejorar la detección en análisis mediante MS/MS de metabolitos en suero humano. Los resultados en cada caso fueron los siguientes:

Capítulo 1

La investigación que se recoge en este capítulo se dedicó a la optimización de un protocolo de tratamiento de muestra previo al análisis de sudor mediante LC–MS/MS. Esta muestra, por la que empieza a aumentar el

interés para estudios clínicos, presenta, como principales ventajas frente a otros biofluidos su muestreo no invasivo (que puede realizarse por personal no especializado), su composición simple y el no requerir estandarización (que sí es necesaria en orina). La utilización de otros biofluidos de mayor complejidad puede llevar a modelos de predicción basados en metabolitos que no son selectivos del proceso biológico en estudio. Esto puede ser debido a interferencias causadas por otros procesos biológicos.

El planteamiento de la optimización implicó el uso de dos modos cromatográficos para el análisis en sudor: El modo en fase reversa utilizando una columna de C18 y el basado en interacción hidrofílica (HILIC) con la columna correspondiente. Respecto a la preparación de la muestra, el análisis directo del sudor permitió detectar 67 y 57 entidades moleculares en el modo de ionización positivo para los métodos analíticos basados en C18 y en HILIC, respectivamente; mientras que en el modo de ionización negativo el número de entidades moleculares fue de 29 y 37, respectivamente, para las columnas citadas. La variabilidad entre individuos fue evaluada mediante análisis de 7 muestras del mismo individuo tomadas en diferentes días y analizadas en duplicado cada día durante un intervalo de 5 días. Este estudio también permitió evaluar la reproducibilidad metodológica de esta etapa, ya que la variabilidad estimada como desviación estándar relativa (RSD) estuvo en el rango del 2 al 21%. Estos valores indican que el sistema de muestreo es suficientemente reproducible para investigación en metabolómica. A pesar del alto número de entidades moleculares detectadas en el modo de ionización positivo, todavía la etapa de preparación de la muestra mejoró la detección de metabolitos. Para ello se compararon diferentes protocolos de preparación de la muestra previos a los dos modos cromatográficos para la obtención del perfil metabolómico global del sudor. Entre los protocolos ensayados, la hidrólisis ácida o básica no resultó recomendable, ya que la muestra se afectaba, tal como puso de manifiesto la aparición de gran cantidad de artefactos. Por el contrario, el análisis directo del sudor después de la dilución apropiada resultó ser una buena opción para obtener una instantánea representativa del metaboloma del sudor. Algunos de los

compuestos exógenos encontrados en el análisis de este biofluido eran componentes del sistema de muestreo. Entre ellos, los más significativos fueron el metil- y el propilparabeno, y especialmente la pilocarpina, que se utiliza para estimular la sudoración antes del muestreo.

Finalmente, la identificación de un alto número de componentes del sudor, en total 43 metabolitos, pertenecientes a una amplia variedad de familias corroboró que el protocolo de análisis basado en LC–QTOF MS/MS es una buena estrategia para analizar sudor. De hecho, se identificaron 19 aminoácidos, incluyendo todos los aminoácidos esenciales, excepto la lisina. También se identificaron compuestos exógenos, como la cafeína y la teofilina, así como ácidos dicarboxílicos, tales como el sebácico, el subérico, el azealico o el butanodioico. Este estudio de identificación contribuye a conocer este biofluido tan poco estudiado y sugiere que, puesto que algunos de los metabolitos identificados habían sido anteriormente evaluados en otros biofluidos como potenciales biomarcadores, el sudor puede también proponerse como biofluido para la búsqueda de biomarcadores.

Capítulo 2

La investigación que conforma este capítulo se dedicó al desarrollo de una estrategia para aumentar la detección de metabolitos. La diversidad química del metaboloma exige nuevas plataformas analíticas para aumentar la capacidad de detección. En este caso, la plataforma estuvo basada en la combinación de un sistema automático de SPE y un equipo LC–QTOF para aprovechar los beneficios de la excelente selectividad de los sorbentes utilizados en SPE y la alta resolución del detector QTOF. El dispositivo de SPE se acopló en línea con el equipo LC–QTOF para la automatización total del método. De esta forma, la intervención del analista se limita a colocar los viales en el automuestreador. La plataforma se aplicó al análisis de suero humano como biofluido modelo. Se comparó la detección de los posibles metabolitos en suero tras el uso de varios sorbentes actuando con mecanismos simples o duales (materiales poliméricos y resinas).

Mediante solapamiento de la capacidad de análisis del equipo SPE–LC–MS/MS con los diferentes sorbentes, se detectaron 3445 entidades moleculares con parámetros predefinidos en términos de carga, distribución isotópica, intensidad de señal y frecuencia de detección en las réplicas analíticas. Una combinación de cuatro sorbentes (resina de polidivinilbenceno, C18 y sorbentes poliméricos con interacciones aniónicas/catiónicas) hicieron posible la detección de más del 81% del total de la detección conseguida con la serie completa de sorbentes ensayados. Por tanto, esta innovación abre una nueva puerta a la mejora de la capacidad de detección en metabolómica utilizando métodos basados en una única plataforma, en este caso LC–MS/MS.

Un desarrollo extra incluido en esta investigación fue el de métodos basados en la combinación de pares de protocolos SPE llevados a cabo en una configuración SPE–SPE–LC–MS. Con este propósito, se hizo circular una única alícuota a través de dos sorbentes SPE complementarios situados en línea. La elución de ambos cartuchos se realizó de forma independiente y secuencial para su análisis mediante LC–MS. Se potenció así la selectividad de este modo de fraccionamiento de la muestra y se maximizó la rapidez del análisis. Las configuraciones en serie diseñadas permitieron conseguir capacidad de detección en el rango del 79.5 al 99.7% del número total de entidades moleculares detectadas mediante análisis separados utilizando la serie completa de sorbentes. Esta alternativa tiene un interés especial en estudios clínicos en los que las muestras son escasas o muy valiosas.

Capítulo 3

Una de las principales debilidades del análisis metabolómico global mediante espectrometría de masas es la necesidad de una información amplia para una identificación correcta de los metabolitos encontrados en un determinado biofluido. Este hecho hace necesaria la inyección del compuesto puro en cuestión, o la obtención de información de alta calidad, tanto de MS como de MS/MS para identificar el compuesto mediante búsqueda en bibliotecas al efecto.

Los métodos de análisis por MS/MS propuestos hasta ahora para identificar las entidades moleculares de interés en un estudio global implican dos o más análisis diferentes. Se requiere una primera inyección de las muestras en el modo MS, seguida de una selección de los iones precursores potenciales de los que se requiere información mediante MS/MS, y un segundo análisis en modo MS/MS orientado. Por otra parte, el modo "auto MS/MS" es una estrategia más corta que consiste en ciclos a lo largo de todo el conjunto de barrido de MS de fragmentación de dos o tres de los iones detectados en el barrido previo. La selección de los potenciales precursores puede restringirse usando rangos de masas o tipo de carga, así como un límite de intensidad de respuesta.

La estrategia GPF consiste en la combinación de diferentes métodos basados en barridos de MS utilizando el mismo rango de masas, pero diferentes intervalos para la selección de iones precursores de manera que se cubra todo el rango del ion precursor deseado. Esta estrategia, aplicada con éxito en proteómica para incrementar el número de proteínas identificadas en una serie de muestras, no había sido utilizada en metabolómica; por tanto, éste fue el objetivo de la investigación recogida en este capítulo.

En metabolómica, la inyección de varias réplicas de la misma muestra es una práctica habitual, ya que la obtención de un perfil metabólico seguro requiere al menos tres réplicas por muestra. Por tanto, el uso de GPF combinado con el número apropiado de réplicas es una herramienta efectiva para aumentar las posibilidades de identificación sin necesidad de realizar posteriores inyecciones de muestra. En esta investigación se ensayaron combinaciones de 2, 3, 4 y 6 intervalos de selección del ión precursor, cubriendo en todos los casos desde 100 a 1000 *m/z*, y se compararon con el método convencional o modo "auto MS/MS".

El biofluido seleccionado para este estudio fue suero, el más comúnmente empleado en análisis metabolómico clínico y nutricional por su papel de principal portador de metabolitos en el organismo humano. El uso de la metodología GPF en metabolómica para el análisis de suero mediante LC–MS/MS ha demostrado ser una estrategia efectiva para incrementar la cantidad de información sobre

algunos compuestos La mejor elección entre todas las combinaciones estudiadas resultó ser un método para dividir el rango de masas de ión precursor en cuatro intervalos. De esta forma se obtuvo información para al menos el 80% de todas las entidades detectadas. Por el contrario, el modo convencional de adquisición de datos ("auto MS/MS") proporcionó información sólo para 48–57% de las entidades moleculares detectadas y, por tanto, fue menos efectivo para la identificación inequívoca de los metabolitos.

Esta investigación ha mostrado que la combinación de diferentes métodos que incluyen distintos intervalos de iones precursores puede incrementar el número de entidades potenciales que proporcionen información MS/MS. Los resultados de esta estrategia se han corroborado mediante identificación de diferentes familias de metabolitos y la evaluación de la información de MS/MS registrada para cada compuesto en cada una de las combinaciones ensayadas.

Sección II. Metabolómica global: Estudios de intervención basados en nutrimetabolómica y búsqueda de biomarcadores de enfermedades

Esta sección se centra en la aplicación del análisis metabolómico global en estudios clínicos para poner de manifiesto la efectividad de esta estrategia. El análisis global en metabolómica puede dividirse en dos categorías genéricas en función del objetivo de la investigación: La primera se refiere a la búsqueda de biomarcadores y la segunda a la comprensión y explicación del efecto de un cambio en una determinada ruta metabólica o un sistema biológico en general. Las dos categorías han sido objeto de la investigación que se recoge en esta sección, que abarca dos diferentes áreas de aplicación ambas orientadas a la medicina personalizada: Diagnóstico, pronóstico y tratamiento de enfermedades, y metabolómica nutricional. En lo que se refiere al área puramente clínica se han considerado dos enfermedades claves: La aterosclerosis, una de las causas principales de muerte en todo el mundo, y el cáncer de pulmón como uno de los tipos de cáncer que causa más muertes al ser el segundo más comúnmente

diagnosticado. En lo que se refiere al área nutricional, la investigación realizada forma parte del proyecto denominado Lipgene, un proyecto europeo para la investigación de la interacción de los nutrientes y el genotipo en el síndrome metabólico, término usado para relacionar varios factores de riesgo en las enfermedades cardiovasculares.

Es conveniente poner de manifiesto que se utilizaron dos tipos de muestras para el desarrollo de la investigación: Suero y sudor, dos biofluidos con muy diferentes características químicas. Se utilizaron también diferentes herramientas estadísticas para la evaluación de los biomarcadores: Generación de paneles formados por combinación de marcadores y estudio de matrices de datos generados en análisis univariante y multivariante.

Se requieren sistemas de masas de alta resolución para el análisis metabolómico global con el fin de conseguir una buena exactitud en la medida de la relación *m/z* que haga posible la identificación del máximo número de metabolitos detectados en el biofluido en estudio. Un equipo QTOF es el preferido en estos casos porque satisface las necesidades del análisis global, ya que permite obtener espectros MS/MS de los iones precursores detectados y proporciona suficiente información para la identificación inequívoca de los metabolitos. Existen bases de datos con información de MS y de MS/MS de una amplia variedad de compuestos. Entre las más importantes de estas bases se encuentra la base de datos de metabolitos basada en MS/MS (Metabolite and Tandem MS Database –METLIN), seguida del banco de masas (MassBank). Otra base de datos como es la del metaboloma humano (Human Metabolome Database – HMDB), no proporciona espectros de MS/MS de alta resolución para todos los metabolitos, ya que la mayoría de los espectros se han adquirido con un equipo de triple cuadrupolo. Otras bases de datos son útiles para obtener información sobre las rutas metabólicas en las que están implicados deter-minados grupos de metabolitos. Éste es el caso de la Enciclopedia Kyoto de Genes y Genomas (Kyoto Encyclopedia of Genes and Genomes —KEGG).

Los resultados obtenidos en la investigación que se recoge en los 4 capítulos que conforman la Sección II de esta Memoria se discuten a continuación.

Capítulo 4

La investigación sobre nutrición es uno de los pilares fundamentales en los que se soportan las vías hacia la Medicina Personalizada. La dieta es uno de los factores externos más importantes que contribuyen a alterar los bloques básicos de la biología de sistemas: El genoma, el transcriptoma, el proteoma y el metaboloma. La relación directa entre el metaboloma y el fenotipo convierte la nutrimetabolómica en un área de gran actividad. En este contexto, el análisis metabolómico global puede ayudar a dilucidar la respuesta biológica en individuos sometidos a dietas de intervención. Éste ha sido el principal objetivo de la investigación que se discute en este capítulo: Comparar las diferencias metabólicas encontradas en los perfiles de suero obtenido de individuos sometidos a 4 dietas de intervención a lo largo de 12 semanas. Las dietas se planificaron como parte del proyecto Lipgene y se diferenciaban en la calidad y cantidad de grasa. Las muestras de suero se dividieron en dos fases mediante extracción líquido–líquido (por tanto, en metabolitos polares y no polares), y cada una de las fases se analizó independientemente mediante LC–QTOF en modo MS/MS.

El análisis estadístico supervisado PLS-DA permitió distinguir diferencias metabólicas asociadas a la dieta de intervención, así como detectar diferencias en los perfiles de metabolitos (polares y no polares) causadas por las dietas ingeridas. Los modelos de discriminación basados en PLS-DA y en análisis de cambio permitieron la identificación de los metabolitos que alteraron su concentración por la dieta de intervención. En relación con la fracción no polar, el papel de los fosfolípidos fue determinante para explicar las diferencias metabólicas entre individuos. Se encontraron diferencias críticas en las concentraciones de las glicerofosfatidilcolinas que incluían PUFAs como sustituyentes.

De hecho, la principal diferencia entre dietas correspondió a tres de los glicerofosfolípidos más comunes en sangre humana —PC(16:0/18:2), PC(18:0/16:2) y PC(18:0/18:2)—, cuyos niveles dependieron estrechamente de la dieta. También los ácidos grasos y las esfingomielinas formaron parte de la lista de metabolitos no polares alterados por las dietas de intervención. Por tanto, la influencia de las dietas en la fracción lipídica fue muy significativa.

Por otra parte, el análisis de la fracción polar también reveló diferencias metabólicas entre individuos en función de las dietas. Las carnitinas, los aminoácidos, los ácidos biliares y los derivados de purina fueron los metabolitos de esta fracción más afectados por la dieta. También se encontró una estrecha conexión entre las principales rutas metabólicas alteradas en los individuos sometidos a los periodos de intervención y el metabolismo de los lípidos. El análisis global permitió establecer diferencias metabólicas asociadas a la dieta, a pesar de la enorme variabilidad biológica observada en la cohorte seleccionada para el estudio.

Capítulos 5 y 6

En estos dos capítulos se recoge el estudio sobre potenciales biomarcadores de aterosclerosis mediante análisis de muestras de suero de pacientes afectados por diferentes enfermedades cardíacas. El estudio que se discute en el Capítulo 5 se orientó a la búsqueda de diferencias metabólicas entre pacientes con angina estable, con infarto de miocardio sin elevación del segmento ST (NSTEMI) o angina inestable, y con infarto agudo de miocardio (AMI). Por su parte, el estudio que recoge el Capítulo 6 se dedicó a la búsqueda de marcadores para el seguimiento de la aterosclerosis mediante comparación de pacientes con angina estable con los que padecían infarto agudo de miocardio. En el primer estudio se ensayaron diferentes modos cromatográficos que permitieron elegir la mejor opción para cubrir el máximo número de entidades moleculares, equivalentes a potenciales metabolitos. De forma adicional, se seleccionó una

etapa simple de precipitación de proteínas como única etapa de preparación de la muestra con el fin de minimizar la pérdida de metabolitos usual de esta etapa.

Una vez establecida la metodología analítica, se analizaron las muestras de suero y los datos generados se trataron de acuerdo con las necesidades de los resultados globales: Alineamiento y extracción de las entidades moleculares y el análisis estadístico correspondiente para encontrar diferencias estadísticas entre los grupos en estudio. Se identificó un grupo de 13 metabolitos que presentó diferencias de concentración significativas entre los tres grupos de pacientes ateroscleróticos. La mayoría de los metabolitos eran lípidos y algunos ácidos biliares, como el cólico y el desoxicólico. La bilirrubina fue el metabolito más diferenciador, con un valor de p de 0.0012. Algunos de los compuestos identificados habían sido relacionados previamente con enfermedades cardio-vasculares o con estrés oxidativo, ambos con enorme influencia en la aterosclerosis. Tras la comparación de las medias de los metabolitos identificados como significativos en los tres grupos, se encontró similaridad entre AMI y angina instable o NSTEMI, mientras que la diferencia entre los valores en pacientes con infarto de miocardio y angina estable fue significativa. Se encontraron también otros metabolitos que, si bien no presentaban una diferencia significativa entre grupos, sí mostraron una tendencia, creciente o decreciente, de angina estable a angina inestable/NSTEMI y a AMI.

La misma plataforma analítica, LC–QTOF MS/MS, se utilizó en el Capítulo 6, ya que había mostrado ser también una estrategia adecuada para la búsqueda de biomarcadores. Este estudio se aplicó a la discriminación de pacientes ateroscleróticos afectados por angina estable y AMI. En este caso se estudió también la influencia de los factores de riesgo (tales como hábito de fumar, obesidad e hipercolesterolemia) mediante análisis de la capacidad de predicción de los potenciales marcadores en pacientes con factor de riesgo específico. Como este segundo estudio se orientó al desarrollo de un panel de marcadores, su evaluación y la de los potenciales marcadores individuales se realizó mediante análisis de curvas ROC y se validó su funcionamiento mediante los tests de diagnóstico médico con clasificación bi-naria. De esta forma, la mejor

capacidad de predicción se obtuvo para los pacientes ateroscleróticos con hipercolesterolemia, para los que la bilirrubina, el 13-HpODE, el 5-HETE y el índice de masa corporal dieron valores de especificidad del 91%, y de sensibilidad entre el 48 y el 62%.

Basado en estos resultados, se ha propuesto un panel de marcadores para la predicción del infarto agudo de miocardio, frente a individuos con angina estable, con una especificidad del 85.1% y una sensibilidad del 80.8%, lo que da lugar a pocas posibilidades de falsos positivos y negativos. Los metabolitos que componen el panel son dos lípidos, el 13-HpODE y el lysoPC(22:6), junto con el índice de masa corporal como variable antro-pométrica relacionada con la obesidad. La confirmación de los resultados del modelo propuesto requiere su aplicación en un estudio a gran escala.

Capítulo 7

El objetivo de la investigación que se discute en este capítulo fue evaluar el potencial de un biofluido menos común como es el sudor para discriminar entre individuos afectados por cáncer de pulmón y donantes sanos. El sudor no ha recibido hasta ahora atención como muestra para la búsqueda de bio-marcadores de cáncer —a pesar de sus comentadas cualidades como el no requerir un muestreo invasivo y tener una matriz menos compleja que la sangre o la orina. Posiblemente esta laguna era debida a la ausencia de un protocolo de muestreo homogéneo y que proporcionara un volumen de muestra suficiente para el análisis. No obstante, los avances en instrumentación analítica han logrado que en la actualidad se disponga de equipos con la sensibilidad adecuada para el análisis de pequeños volúmenes de muestra. Este hecho, junto con la mejora de los protocolos de muestreo del sudor, suficientemente reproducibles para estudios de cáncer de pulmón, han hecho posible el uso de este biofluido para un potencial diagnóstico de esta enfermedad. Esta búsqueda responde a la necesidad de nuevas herramientas para el diagnóstico de este cáncer, ya que los

tests que existen en la actualidad son invasivos y tienen un coste alto, lo que hace impracticable su aplicación a todos los individuos con riesgo de padecerlo.

Como en capítulos anteriores, se utilizó la plataforma LC–QTOF MS/MS junto con el modo más simple de preparación de la muestra, que consistió en dilución con la fase móvil cromatográfica inicial. Con estas premisas, el potencial del sudor como biofluido para su implantación en diagnóstico de cáncer se demostró de la siguiente forma: Se construyó un modelo de predicción basado en un panel de metabolitos que incluyó aminoácidos, azúcares y algunos lípidos y que permitió discriminar entre pacientes con cáncer y un grupo control. Su alto valor predictivo permite prever su potencial uso para reducir el número de posibles casos que tienen que someterse a tests confirmatorios.

La capacidad de discriminación de metabolitos individuales reveló que un trisacárido fosfato presentaba el valor más alto de sensibilidad para una especificidad mínima del 80%. Sin embargo, este compuesto no estaba presente en los dos paneles de tres metabolitos generados para la predicción de cáncer de pulmón. Los dos paneles incluían maltotriosa y ácido nonanedioico, en combinación con γ-GluLeu y MG(22:2). Ambos paneles mejoraban de forma significativa la capacidad de discriminación de los metabolites independientes. El panel se caracterizó por el 100% de especificidad y el 63.6% de sensibilidad; por tanto, la presencia de falsos negativos fue 0%. El segundo panel proporcionó valores de especificidad y sensibilidad del 82%.

Estos resultados preliminares hacen muy conveniente un estudio a gran escala para validar los paneles propuestos con un doble objetivo: Reducir el número de individuos que tengan que someterse a un test confirmatorio y detectar el cáncer de pulmón en una etapa lo más temprana posible.

Sección III. Análisis orientado de potenciales biomarcadores

El análisis orientado en metabolómica se define como aquél que se aplica a una muestra para la cuantificación de un grupo o familia de compuestos, o

incluso un único metabolito. A pesar de que sólo se pueden detectar compuestos predefinidos utilizando esta estrategia metabolómica, posee una serie de ventajas respecto al análisis global. La principal de ellas es que tanto la preparación de la muestra como la detección se optimizan para reducir el número de interferentes y que la selectividad, y también la sensibilidad, sean máximas con la plataforma seleccionada para el análisis. Con este criterio, en el caso de determinar compuestos no polares, se prefiere una columna de C18 o similar, mientras que para compuestos polares la mejor selección es una HILIC o de intercambio iónico. La preparación de la muestra depende, lógicamente, de la naturaleza de los analitos y en algunos casos se requiere una etapa de derivatización, como ocurre en el análisis mediante GC–MS. En otros casos la extracción en fase sólida es la major alternativa para eliminar los interferentes y preconcentrar los metabolitos problema.

El número de compuestos considerados en cada estudio puede variar desde un único metabolito a una familia que comprenda un número grande de compuestos, como es el caso de los fosfolípidos. Cuando el número de compuestos es suficientemente pequeño para optimizar la preparación de la muestra y la detección para cada uno de ellos, y se puede obtener su recta de calibrado individual —porque la existencia de patrones lo permite— se puede conseguir la cuantificación de todos los compuestos problema. Por el contrario, si el número de compuestos es grande y no existen patrones para todos ellos —y, por tanto, la obtención de las curvas de calibrado no es completa— la cuantificación de los metabolitos que carecen de patrones se hace generalmente de forma relativa con respecto al metabolito para el que existe patrón y que tiene la estructura más similar a él.

Los paneles de compuestos seleccionados en esta investigación para análisis orientado estuvieron compuestos por metabolitos implicados en rutas metabólicas conocidas y que juegan un papel relevante en ciertas patologías. Además, estos metabolitos habían sido descritos previamente como potenciales biomarcadores. Tres de las cuatro plataformas analíticas optimizadas en esta sección se orientaron al estudio de diferentes familias de metabolitos, como

aminoácidos esenciales, compuestos implicados en el ciclo de los ácidos tricarboxílicos y fosfolípidos. La cohorte seleccionada para el estudio estaba formada por pacientes diagnosticados con diversas afecciones cardíacas. Con la cuarta de las plataformas optimizadas se analizó un péptido antimicrobiano del que se determinó comparativamente su concentración en suero de individuos sanos y de pacientes de la unidad de cuidados intensivos.

En relación a la naturaleza química de los compuestos analizados, dos de los métodos se basaron en LC–MS, uno en GC–MS y otro en infusión directa en MS/MS. Tanto en este último método como en los basados en LC–MS, se aprovecharon las ventajas que aporta la SPE como preparación de muestra de forma automatizada y con acoplamiento en línea a etapas posteriores del proceso analítico.

En lo que se refiere a las estrategias de tratamiento de datos, se utilizó un esquema general similar al de la sección previa —análisis univariante y multivariante— junto con nuevas herramientas con diferente capacidad. Como en la sección anterior, se utilizaron las dos formas de análisis univariante, paramétrico y no paramétrico, pero también estrategias de análisis multivariante supervisado y no supervisado. También se utilizaron diseños multivariantes y univariantes para obtener el valor óptimo de cada parámetro que afectaba al método en cuestión, siempre en función de la metodología a desarrollar.

Capítulo 8

Este capítulo está dedicado a una plataforma para el análisis de fosfolípidos en suero con el propósito de determinar el perfil de estos compuestos en pacientes ateroscleróticos. Los fosfolípidos son una familia de lípidos que, además de ser componentes mayoritarios de las membranas celulares, constituyen una parte importante de la composición lipídica de las lipoproteínas presentes en la sangre. La mayoría de los fosfolípidos contiene un diglicérido o monoglicérido, un grupo fosfato y una molécula orgánica simple, tal como colina, etanolamina o serina, que dan nombre a las diferentes clases de fosfolípidos. Una

excepción a esta regla es la esfingomielina, que deriva de la esfingosina en lugar de derivar del glicerol. Como esta familia de lípidos abarca un gran número de metabolitos que incluye glicerofosfatidilcolinas (PC), glicerofosfatidiletanolaminas (PE), glicerofosfatidilserinas (PS) y esfingomielinas (SM), entre otras, es parti-cularmente difícil construir las curvas de calibración para todos ellos. Por esta razón, los perfiles de fosfolípidos del suero se obtuvieron mediante LC–QTOF MS/MS, una plataforma que permite la identificación de fosfolípidos gracias a su alta resolución. La identificación de fosfolípidos en suero puede llevarse a cabo por la información que proporcionan en MS/MS, ya que su modelo de fragmentación es conocido. Se combinaron, además, los modos de ionización positivo y negativo para obtener resultados complementarios.

En cuanto a la preparación de la muestra, se centró en el aislamiento de los fosfolípidos de la matriz del suero y a su preconcentración. Con estos fines, una etapa de extracción en fase sólida utilizando cartuchos de un sorbente que interacciona selectivamente con el grupo fosfato (sílice recubierta de circonio) permitió una retención selectiva de los metabolitos con este grupo funcional. Ochenta y un fosfolípidos se detectaron e identificaron utilizando esta plataforma de los que 48 estaban presentes al menos en el 75% de las muestras estudiadas (140 de pacientes diagnosticados con aterosclerosis coronaria, 72 con angina inestable o NSTEMI y 68 con angina estable).

El análisis estadístico de los datos obtenidos permitió identificar cuatro fosfolípidos a concentraciones significativamente diferentes en pacientes con angina estable con respecto a los que padecían angina inestable o NSTEMI: LysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) y SM(d18:2/14:0).

Se utilizó una herramienta analítica conocida como ROCCET para construir un panel de marcadores con los anteriores metabolitos con el que discriminar entre pacientes ateroscleróticos que padecían angina estable o inestable/NSTEMI. El panel presentó una capacidad de predicción de la enfermedad en el set de entrenamiento del 70.7 y 66.0% para angina estable e inestable/NSTEMI, respectivamente. Además, la curva ROC del modelo presentó

un área bajo la curva de 0.715 para el set de entrenamiento y de 0.747 para el set de validación externa, mostrando un 66 y 70% de sensibilidad y especificidad, respectivamente. Estos resultados merecen un estudio a gran escala para validar el modelo propuesto y asegurar la utilidad de la estrategia basada en LC–QTOF MS/MS para establecer el perfil de fosfolípidos que permita el desarrollo de herramientas para asistir en diagnóstico clínico.

Capítulo 9

Considerando el perfil de aminoácidos esenciales y su relevancia biológica, el objetivo del presente capítulo fue desarrollar una plataforma automatizada para la determinación de aminoácidos esenciales en suero con vistas a conocer si existen diferencias significativas entre sus niveles dependiendo del tipo de pacientes ateroscleróticos. Con este propósito, se adoptó el sistema SPE–LC–MS/MS como configuración instrumental.

El número de metabolitos en este caso permitió la optimización completa del proceso analítico, así como la caracterización del método desarrollado: Curvas de calibración, rango dinámico, límites de detección y de cuantificación, reproducibilidad y repetitividad. Se tuvieron en cuenta las características físico-químicas de los analitos —aminoácidos— para el diseño del método. Con este criterio, en la optimización de la preparación de la muestra mediante SPE se utilizaron cartuchos catiónicos, dado que los aminoácidos adquieren carga positiva en medio ácido. Teniendo en cuenta el carácter polar de los metabolitos, en la etapa de separación cromatográfica, se seleccionó una columna HILIC (fase móvil acuosa, inicial con 90% de acetonitrilo y final con 40%, en todo momento 5 mM en formiato amónico y con pH 6). Puesto que las fases cromatográficas no coincidían con el disolvente requerido para eluir los aminoácidos del cartucho (90% acetonitrilo, 5% amoníaco, 5% agua), se eligió el modo de elución "focusing". Este modo está especialmente indicado para la elución con un mínimo volumen de eluyente diferente de la fase móvil. El caudal de la etapa de

elución debe ser pequeño para minimizar su influencia en la separación cromatográfica.

Una vez optimizado, el método se aplicó al análisis de muestras de suero de 122 pacientes ateroscleróticos (80 con angina estable y 42 que habían sufrido previamente un infarto agudo de miocardio). El análisis de esta serie de muestras reveló que la isquemia, presente en los pacientes que habían sufrido un infarto, tenía un efecto significativo en la concentración de dos aminoácidos esenciales: Treonina y lisina. Este efecto era más pronunciado en mujeres que en hombres. La influencia de la obesidad, un factor de riesgo clave en enfermedades cardiovasculares, también se elucidó poniéndose de manifiesto niveles de valina y metionina más altos en suero de individuos obesos. Esta investigación permitió establecer diferencias significativas en el perfil de aminoácidos esenciales presentes en suero de individuos diagnosticados con aterosclerosis, con especial énfasis en la importancia de factores de riesgo para mejorar la discriminación.

Capítulo 10

En este capítulo se discuten la puesta a punto y la aplicación de un método para la determinación en suero de compuestos implicados en el ciclo de los ácidos tricarboxílicos (TCA) mediante el uso de GC–MS tras una etapa de derivatización, una estrategia muy común para la determinación de esta familia de compuestos. El ciclo de los TCAs también se conoce como ciclo del ácido cítrico o ciclo de Krebs', y participa en la generación de energía a partir de carbohidratos, grasas y proteínas. Está también relacionado con la generación de ciertos aminoácidos y ácidos grasos, y es la ruta metabólica central de los organismos aeróbicos.

Se optimizaron tanto la preparación de la muestra como la determinación para conseguir la máxima sensibilidad y selectividad posibles. El método se basó en, (i) precipitación de proteínas con metanol, (ii) limpieza de la fase líquida mediante extracción líquido–líquido con cloroformo para eliminar los interferentes no polares, (iii) derivatización por sililación; (iv) análisis mediante GC–

MS. El método se aplicó a muestras de suero de 223 pacientes diagnosticados con diferentes enfermedades cardiovasculares (172 con lesiones coronarias significativas y 51 afectados por otros problemas cardiovasculares, pero sin lesiones coronarias). Los compuestos del TCA son primariamente metabolitos que pueden verse afectados por un número amplio de factores internos y externos. Por esta razón es importante considerar la influencia de los factores de riesgo cuando se proponen los metabolitos del ciclo TCA como potenciales biomarcadores en diagnosis clínica. Por tanto, se estudió la influencia de tres factores de riesgo como la hipercolesterolemia, la obesidad y el hábito de fumar en los niveles en suero de los metabolitos del TCA en paciente diagnosticados con lesiones coronarias. La consideración de estos factores de riesgo resultó crucial para explicar la variabilidad de esos metabolitos en pacientes con lesión coronaria frente a los individuos control. En este estudio, el análisis estadístico estuvo soportado en curvas ROC y sus parámetros característicos para validar la capacidad de discriminación de los metabolitos del TCA.

Los resultados obtenidos revelaron que los ácidos málico y α-cetoglutárico conducen a modelos con alta especificidad y aceptable sensibilidad para la discriminación entre pacientes con o sin lesiones coronarias considerando también la obesidad (un factor de riesgo que influyó en los niveles de estos dos metabolitos). No obstante, el modelo de discriminación del ácido α-cetoglutárico fue menos robusto que el del ácido málico.

Capítulo 11

La plataforma analítica orientada a la que se dedica este capítulo se centró en la cuantificación de péptidos de interés clínico. La plataforma se fundamenta en el acoplamiento en línea de la SPE y la espectrometría de masas en tándem con infusión directa, con el fin de hacer posible el desarrollo de métodos rápidos, con alta sensibilidad y selectividad. La aplicabilidad de la plataforma se comprobó mediante su aplicación a la determinación de un péptido antimicrobiano en suero humano: La catelicidina. El modo de seguimiento

selectivo de reacciones (SRM) se configuró con tres transiciones confirmadas por fragmentación *in silico* del péptido en cuestión. Se evitó el uso de un estándar interno mediante validación de la SPE realizando ensayos con configuraciones de doble cartucho para asegurar la retención cuantitativa de la catelicidina.

Mediante el sistema automatizado de SPE se llevó a cabo la desalinización y desproteinización de la muestra de forma eficaz utilizando como sorbente una resina polimérica de divinilbenceno. La eficacia de esta etapa es de crucial importancia en análisis de biofluidos, en los que frecuentemente la etapa de preparación de la muestra incluye una etapa de desproteinización con disolventes orgánicos. El proceso analítico requirió sólo 12 min por muestra. Los límites de detección y de cuantificación fueron 2.5 y 8.25 μg/L, respectivamente (0.20 and 0.66 pg en columna), que ponen de manifiesto la alta eficiencia de la combinación SPE–MS/MS. La repetibilidad y la reproducibilidad en el laboratorio fueron 2.4% y 2.7%, respectivamente; lo que demuestra la precisión del método desarrollado.

Como la catelicidina participa en la respuesta inmunitaria, operando en primera línea de los mecanismos de defensa, el método se validó por aplicación a muestras de suero de pacientes de la unidad de cuidados intensivos (n=17) y de donadores sanos (n=23). Aunque los niveles de catelicidina en los pacientes en cuidados intensivos fueron mayores que en los inviduos sanos, no se encontraron diferencias significativas entre ambos grupos.

La plataforma analítica propuesta en este capítulo puede ser de interés para el análisis de otros péptidos, tales como biomarcadores clínicos y de proteínas problema después de hidrólisis enzimática para generar péptidos representativos. Por tanto, con cambios adecuados o sin ellos, podría utilizarse para cualquier otro péptido con una disminución considerable del coste por análisis, si se compara con el de los tests basados en inmunoensayo. También se consigue una completa automatización sin necesidad de usar estaciones de trabajo robotizadas, así como un grado extra de selectividad al suprimirse las reacciones cruzadas.

DISCUSSION OF THE RESULTS

The current regulation of the University of Córdoba on the writing of doctoral dissertations dictates that any articles deriving from the doctoral work, whether published or awaiting publication, are to be included in the Thesis report as such and that the report should include a section presenting a joint discussion of the results.

The research leading to the main part of this doctoral work revolved largely around clinical metabolomics. This choice of research topic was prompted by the strong impact of this omics discipline on two especially attractive fields, namely: (*a*) diagnosis, prognosis and treatment of diseases; and (*b*) nutrition. These two fields are the cornerstones of PM. The research presented here is described in three different sections according to the aim and the specific metabolomic strategy used for each purpose.

Thus, Section I focuses on the development of new methodologies to improve the basic analytical properties sensitivity and selectivity, and also to increase metabolite coverage and facilitate identification of potential biomarkers.

The second and third sections are devoted to two basic metabolomics strategies: untargeted and targeted analysis. Specifically, Section II deals with profiling analysis, which is the most widely used mode of untargeted analysis in clinical metabolomics and aims to cover the metabolomes of specific biofluids or tissues as comprehensively as possible.

By contrast, Section III is concerned with the determination of specific groups or families of compounds which are of clinical interest because of their involvement in crucial biological pathways that can be altered or deregulated by certain pathological conditions. The target metabolites addressed in this section include essential amino acids, metabolites involved in the tricarboxylic acids cycle, phospholipids and the antimicrobial peptide cathelicidin.

The results pertaining to each section are summarized below. One common link between the three is the use of mass spectrometry for detection. The analytical sensitivity, selectivity, accuracy, precision and resolution of mass spectrometry make this technique the most suitable tool for targeted and untargeted metabolomic analysis of clinical samples. Both metabolomics strategies have taken advantage of this technique for qualitative and quantitative analyses. Thus, untargeted analysis has benefited greatly from the high resolution (i.e., high *m/z* accuracy), scan velocity and sensitivity, and competitive purchasing cost, of QTOF mass spectrometers, which have turned it into the preferred choice for qualitative analysis and relative quantitation. On the other hand, QqQ mass spectrometers are better suited to targeted analysis by virtue of their high sensitivity and selectivity for confirmatory analysis —and hence for quantitation. Given the complexity of clinical samples, mass spectrometers are better suited to clinical metabolomics, especially when used in combination with chromatographs. The choice of a specific technique (gas chromatography or liquid chromatography, mainly) for separation, when needed, depends on the chemical properties of the target metabolites.

Section I. Methodological development and innovation in metabolomic analysis

Metabolomics analysis continues to be subject to some shortcomings for untargeted and profiling analysis despite recent cutting-edge technological advances including accurate, high resolution mass spectrometry measurements for expeditious detection of metabolites.

One of the problems with profiling metabolomics is the difficulty of covering all possible metabolites with a single analytical platform; this is a result of metabolomes encompassing a wide variety of metabolites differing widely in chemical structure and concentration. One other crucial problem arises from the complexity of some biofluids typically used in clinical analysis (viz., serum,

plasma, urine, saliva), which further hinders development of clinical metabolomics applications.

The previous problems were addressed in three different ways here. One involved preparing samples of such a scarcely studied biofluid as sweat, which has substantial advantages including non-invasive sampling and a simple composition relative to conventional biofluids such as blood or urine. Using new, alternative types of samples in clinical analysis can be useful to overcome the low selectivity of conventional biofluids to metabolic changes associated to physiological changes. One other challenge was developing an effective sample preparation strategy to increase metabolite coverage in the LC–MS/MS analysis of clinical samples. The strategy used for this purpose was an automated SPE system coupled on-line to LC–QTOF. The SPE system was used with two different sorbents possessing complementary retention properties for sequential elution to the chromatograph. The third challenge was another methodological innovation, namely: using gas phase fractionation in metabolomics to improve metabolite detection in the MS/MS analysis of human serum. The results provided by each approach are summarized below.

Chapter 1

The research work described in this chapter was aimed at developing an optimized sample treatment protocol for the LC–MS/MS analysis of sweat, an uncommon biofluid which is arising increasing interest for clinical studies by virtue of its easy, non-invasive sampling —even by unskilled personnel—, simple composition and, unlike urine, the need for no standardization. More complex biofluids have led to prediction models based on unselective metabolites by effect of interferences from other biological processes.

The experimental optimization plan involved using two different chromatographic modes for sweat analysis, namely: a reversed phase mode using a C18 column and a HILIC mode using an appropriate column. Direct analysis of sweat enabled detection of 67 and 57 molecular features with the C18 and HILIC

analytical method, respectively, in the positive ionization mode, and 29 and 37 features, respectively, in the negative ionization mode. Intra-individual variability was assessed by duplicate analysis of 7 samples from the same individual on different days over a period of five. This study also allowed evaluation of the methodological reproducibility of this step, the estimated variability as RSD of which ranged from 2 to 21%. Therefore, the sampling system was reproducible enough for metabolomic research.

Despite the large number of potential molecular features detected in the positive ionization mode, metabolite coverage was further improved by the sample preparation step. To this end, various preparation protocols were used in the two chromatographic modes for the overall metabolomics profiling analysis of sweat. Sweat hydrolysis, whether acid or basic, proved unsuitable for metabolite profiling because it altered the integrity of the samples judging by the presence of many artifacts. By contrast, direct analysis after dilution proved effective to obtain a representative snapshot of sweat metabolome. Also, C18 SpinColumn SPE cartridges resulted in improved sensitivity and reduced the presence of exogenous compounds. Some exogenous compounds detected in this biofluid were components of the sweat extraction system (particularly methylparaben, propylparaben and pilocarpine, the last of which was used to stimulate sweating prior to sampling). A clean-up step with HILIC SPE cartridges was found to be preferable for the analysis of lipid metabolites such as monoacylglycerols.

The optimized LC–QTOF MS/MS protocol afforded the identification of a large number of sweat components (a total 43 metabolites from a wide variety of families), which testifies to its effectiveness for analysing sweat. The metabolites identified included all essential amino acids except lysine —a total of 19 compounds. Exogenous compounds such as caffeine and theophylline were also identified, and so were various dicarboxylic acids (sebacic, suberic, azelaic and butanedioic). The results of this identification study can help us better understand this poorly known biofluid and suggest that, since some of the metabolites

identified had previously been assessed as potential biomarkers for other biofluids, sweat may be a candidate biofluid for finding new biomarkers.

Chapter 2

The work described in this chapter focused on the development of an innovative approach to increase metabolite coverage. The high chemical diversity of the metabolome requires new analytical platforms to improve existing detection capabilities. The platform used for this purpose here was a combination of an automated SPE device and an LC–QTOF system intended to take advantage of the high selectivity of SPE sorbents and the high resolution of QTOF detectors. The SPE device was coupled on-line to the LC–MS/MS instrument for fully automated operation —in fact, all the analyst needs to do is placing sample vials on the autosampler.

The proposed approach was used for the analysis of human serum as a model biofluid. Detection of tentative metabolites in serum with various sorbents acting via single or dual retention mechanisms was compared, using polymers and resins for dual retention. Using different sorbents allowed the detection capabilities of SPE–LC–MS/MS to be maximized and up to 3445 molecular entities detected under specific conditions of charge state, isotopic distribution, signal intensity and detection frequency. A combination of four sorbents (viz., polydivinylbenzene resin, C18, and anionic and cationic mixed mode materials) afforded more than 81% of the total coverage obtained with the whole set. Therefore, the proposed approach opens a new door to enhancing the detection capabilities of metabolomic methods using a single platform (LC–MS/MS here).

An additional study was conducted to develop other methods using combinations of two SPE protocols conducted on automated SPE–SPE–LC–MS configurations. For this purpose, a single aliquot was serially circulated through two complementary SPE sorbents, the two cartridges being independently eluted and the eluted metabolites sequentially analysed by LC–MS. This increased the selectivity of this hyphenated sample fractionation scheme and maximized

throughput. Thus, the serial configurations used afforded detection coverages ranging from 79.5 to 99.7% of all molecular features detected by separate analysis with the whole set of sorbents. This approach can be very interesting for clinical studies on scant or highly valuable samples.

Chapter 3

One of the weaknesses of untargeted metabolomic analysis by mass spectrometry is the need to obtain a large amount information for correct identification of metabolites in the particular biofluid under study. This requires either separate injection of the target compounds in pure form or acquisition of high-quality MS and MS/MS information to identify them through a library search.

Existing MS/MS methodologies for identifying molecular features in untargeted analysis involve two or more steps including injection of samples in the MS mode, selection of potential precursors for obtainment of MS/MS information and targeted MS/MS analysis. The "auto MS/MS" mode is comparatively less time-consuming as it involves cycling through the chromatographic run of MS scans and subsequent fragmentation of two or three of the ions detected in the previous scan. Selection of potential precursors can be restricted to specific mass ranges or charge states, and an abundance threshold be established.

The strategy based on GPF uses a combination of MS scans over the same mass range but different intervals to select precursor ions spanning the whole desired precursor range. This strategy, which had been successfully applied to increase the number of proteins identified in proteomics sample batches, had never before been used in metabolomics. This was thus the primary goal of the research work described in this chapter.

Because accurate metabolomics profiling requires using at least three replicates per sample, replication is a usual practice in metabolomics studies. Using GPF in combination with an adequate number of replicates is an effective

approach to enhancing identification capabilities without the need for further sample injections. In this doctoral work, we used combinations of 2, 3, 4 and 6 intervals to select precursors spanning the *m/z* range from 100 to 1000 for comparison with the results of the conventional "auto MS/MS" mode.

This study was conducted on serum, which is the most widely used biofluid for clinical and nutritional metabolomic analyses by virtue of its role as a primary carrier of metabolites in the human body. Using GPF methodology for serum analysis by LC–MS/MS in metabolomics has proved an effective strategy for increasing the amount of MS/MS information available on some compounds. A combined method splitting the range of precursor ion masses into four intervals was found to be the best choice; thus, it provided MS/MS information for at least 80% of all detected entities. By contrast, the conventional "auto MS/MS" data acquisition mode afforded identification of only 48–57% of all detected molecular entities and was therefore less effective for unequivocal identification of metabolites.

Based on the results, a combination of methods using different intervals of precursor ion masses can increase the number of entities potentially detected from MS/MS information, as confirmed by identifying various families of metabolites and assessing recorded MS/MS information for each target compound in each combination.

Section II. Untargeted metabolomics: intervention study based on nutrimetabolomics, and search for disease biomarkers

This section focuses on the use of different untargeted metabolomic approaches to confirm the effectiveness of this strategy for clinical studies. Untargeted metabolomics studies can be classified into two general categories depending on the particular purpose, namely: (*a*) discovery of biomarkers, and (*b*) understanding a specific metabolic pathway or an overall biological system.

Both categories were addressed here in relation to two aspects of Personalized Medicine, namely: (*a*) diagnosis, prognosis and treatment of diseases, and (*b*) nutritional metabolomics.

The purely clinical area was examined in connection with atherosclerosis, which is one of the leading cause of death worldwide, and lung cancer, which is the leading and second most frequently diagnosed cause of cancer death. The nutritional area was explored in research conducted within the framework of the Lipgene study. Lipgene is a European project focusing on the interaction of nutrients and genotype in the metabolic syndrome —a term that encompasses several risk factors for cardiovascular disease. Experiments were conducted on two different types of matrix: serum and sweat, which are two biofluids of different chemical diversity. Various statistical tools for evaluation of biomarkers, panels consisting of several markers, and univariate and multivariate analysis of the data set, were used for this purpose.

Untargeted metabolomics analyses require using high mass resolution systems in order to ensure a high enough *m/z* accuracy to identify as many metabolites in the target biofluid as possible. QTOF is a preferential tool here as it fulfils the requirements of untargeted analysis while facilitating acquisition of MS/MS spectra for precursor ions that contain enough information for the unequivocal identification of metabolites. A number of MS and MS/MS databases containing useful information for a wide variety of compounds exist, especially prominent among which are the Metabolite and Tandem MS Database (METLIN) and the MassBank. By contrast, the Human Metabolome Database (HMDB) contains MS/MS spectra with inadequate resolution for some metabolites because most of the spectra were acquired with a triple quadrupole system. Some databases are useful to obtain information about metabolic pathways involving a specific group of metabolites; such is the case with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

The results obtained in the work described in the four chapters of Section II are discussed in detail below.

Chapter 4

Nutritional research provides another crucial pillar for Personalized Medicine. Diet is one of the most important external factors contributing to changes in the main blocks of the Systems Biology concept, namely: genome, transcriptome, proteome and metabolome. The direct relationship between metabolome and phenotype has turned nutrimetabolomics into a very active research area. For example, untargeted metabolomic analysis can be useful to elucidate the biological response of individuals under intervention diets. This was the main purpose of the research work conducted here: to compare the metabolic differences between serum profiles from individuals following four different intervention diets for 12 weeks. The diets were planned as part of the Lipgene project and differed in fat quality and quantity. Liquid–liquid extraction was used to split serum samples into two fractions (polar and non-polar metabolites) that were analysed separately by LC–QTOF in the MS/MS mode.

Supervised statistical analysis (PLS-DA) allowed metabolic differences associated to the intervention diet, and differences in metabolite profiles (polar and non-polar compounds) due to each diet, to be detected. Discrimination models based on PLS-DA and fold change analysis led to identification of metabolites whose concentration in serum was altered by the intervention diet. Phospholipids in the non-polar fraction were found to play a central role in metabolic differences between individuals. Thus, critical differences in the concentrations of glycerophosphatidylcholines including PUFAs as substituents were observed. The greatest differences between diets were those in three of the most common glycerophospholipids in human blood [PC(16:0/18:2), PC(18:0/16:2) and PC(18:0/18:2)], the levels of which differed between diets. Non-polar metabolites altered by the intervention diets included some essential fatty acids and sphingomyelins. Therefore, the diets had a considerable influence on the lipid fraction.

Analysis of the polar fraction also revealed metabolic differences between individuals under different diets. Carnitines, amino acids, bile acids and purine base derivatives were among the most representative metabolites in this altered fraction. A strong connection between the major pathways altered —most relating to lipid metabolism— in individuals under the intervention diets was observed. Untargeted analysis allowed diet-related metabolic differences to be identified despite the high biological variability in the studied cohort.

Chapters 5 and 6

These two chapters describe the metabolomic study of potential biomarkers for atherosclerosis identified by analyzing serum samples from patients suffering from various heart diseases. The study described in Chapter 5 involved a search for metabolomic differences between patients with stable angina, non-ST elevation myocardial infarction (NSTEMI) or unstable angina and acute myocardial infarction (AMI), whereas that described in Chapter 6 was aimed at finding effective markers for monitoring atherosclerosis by comparing patients with stable angina and acute myocardial infarction. In the former study, various chromatographic modes were used to identify that leading to the greatest possible coverage of potential molecular features —which are equivalent to tentative metabolites. The only sample pretreatment needed was protein precipitation, which minimized metabolite losses during the sample preparation step.

The ensuing methodology was used to analyze serum samples and the data thus obtained were processed according to the particular needs for untargeted results, namely: alignment and extraction of potential molecular features, and statistical analysis for differences between groups. A group of 13 metabolites with significant differences in concentration between the three different groups of atherosclerotic patients was identified. The metabolites were mostly lipids, but also included some bile acids (e.g., cholic and deoxycholic acids). Bilirubin was the most discriminating metabolite, with $p = 0.0012$. Some of the com-pounds identified had previously been associated to cardiovascular disease

or oxidative stress, both of which are highly influential on atherosclerosis. A comparison of means for significantly identified metabolites between the three groups revealed some similarity between AMI and unstable angina/NSTEMI, and the greatest differences in metabolites levels to be those between myocardial infarction and stable angina patients. Some other metabolites were not significantly different between groups, but tended to increase or decrease from stable angina to unstable angina/NSTEMI to AMI.

The LC–QTOF MS/MS analytical platform was also used in the research work described in Chapter 6 and found to be a suitable choice for biomarker discovery. This study was applied to the discrimination of atherosclerotic patients affected by stable angina and by a previous AMI episode. The influence of risk factors such as smoking habit, obesity and hypercholesterolemia was examined by assessing the predictive ability of potential markers in patients with a specific risk factor. Because the aim was to develop a panel of markers, assessment was based on ROC curve analysis, which is the golden standard for assessing performance of medical diagnostic tests with binary classification. The best predictions were those for atherosclerotic patients with hypercholesterolemia as obtained with bilirubin, 13-HpODE, body mass index (BMI) and 5-HETE, which had 91% specificity and 48–62% sensitivity.

A panel of markers for predicting acute myocardial infarction versus stable angina was compiled that featured 85.1% specificity and 80.8% sensitivity, and hence few false negatives or positives. The metabolite panel comprised the lipids 13-HpODE and lysoPC(22:6) in addition to BMI, which is an obesity-related anthropometric variable. In any case, these results and the proposed model require validation in a large-scale study.

Chapter 7

The aim of the work described in this chapter was to evaluate the potential of such a scarcely studied biofluid as sweat for discriminating individuals with lung cancer from healthy donors. Sweat has in fact been scarcely

used as a sample to search for cancer biomarkers despite its above-described advantages including non-invasive collection and lesser complexity than blood or urine. This has largely been the result of the lack of uniform sweat sampling protocols providing large enough volumes for analysis. However, advances in analytical instrumentation have raised sensitivity to a level facilitating the analysis of small sample volumes. This fact, together with improvements in sweat collection protocols which have made them reproducible enough for biomarker searching, has turned sweat into a useful sample for diagnosis of lung cancer. New tools for diagnosing lung cancer are needed, however, because currently available tests are invasive and expensive —and hence impossible to apply to all individuals at risk.

As in previous chapters, LC–QTOF MS/MS was used in combination with the most simple possible sample treatment: dilution with the initial chromatographic phase. This analytical platform was used to assess the potential of sweat as a biofluid for diagnosing lung cancer by developing a prediction model based on a panel of metabolites including amino acids, sugars and some lipids to discriminate patients with lung cancer from a control group. The high negative predictive value of this approach testifies to the potential usefulness of this biofluid for reducing the number of positives requiring confirmatory analysis.

A trisaccharide phosphate was found to be the best discriminator among individual metabolites. However, this compound was not present in both three-metabolite panels used to predict lung cancer. The two panels included maltotriose and nonanedioic acid, in combination with γ-GluLeu and MG(22:2). Both significantly improved the discrimination capabilities of individual metabolites. Thus, the former panel featured 100% specificity and 63.6% sensitivity —and hence a false negative rate of 0%—, and the latter 81.0% specificity and sensitivity.

These preliminary results exposed the need for a large-scale study to validate the proposed panels in order to reduce the proportion of individuals requiring confirmatory testing, and to be able to detect lung cancer at earlier stages.

Section III. Targeted analysis of potential biomarkers

Targeted analysis in metabolomics is defined as the analysis of a sample to quantify a specific group or family of metabolites, or even a single metabolite. Although it only allows certain compounds to be detected, this metabolomic strategy has some clear-cut advantages over untargeted analysis. The main advantage is that sample preparation and detection can be optimized to reduce the number of interferents and maximize the selectivity and sensitivity by using an appropriate analytical platform. Thus, a C18 column is to be preferred for non-polar compounds and a HILIC column for polar compounds. Sample treatment is also analyte-dependent and occasionally includes a derivatization step (e.g., with GC–MS analysis) or requires SPE for removal of interferents and preconcentration of target metabolites.

The number of compounds included in each targeted study can range from a single metabolite to a large family (e.g., phospholipids). When the number of target compounds is small enough to allow both sample preparation and detection to be optimized, and a calibration curve for each individual metabolite obtained, absolute quantitation of the target compounds is feasible. On the other hand, if the number of target compounds is too large or some lacks a standard from which a calibration curve can be constructed, then absolute quantitation of all metabolites is impossible and those lacking a standard must be quantified in relation to the most similar metabolite having one.

The panel of compounds used here for targeted analysis consisted of metabolites involved in some well-known metabolic pathways that play a central role in certain medical conditions. The metabolites concerned had been previously studied as potential biomarkers. Three of the four platforms used were applied to different metabolite families including essential amino acids, compounds involved in the tricarboxylic acids cycle and phospholipids. The target

cohort consisted of patients diagnosed with diverse cardiovascular diseases. The fourth platform used an antimicrobial peptide as target to compare serum levels in healthy individuals and intensive care patients.

Based on their chemical nature, the target metabolites were determined with two LC–MS methods, a GC–MS method and a direct-infusion MS/MS method. The last one, and those using LC–MS, benefited from the advantage of SPE for automated sample preparation coupled on-line with other steps of the analytical process.

Data processing was based on the general scheme used in the previous section (univariate and multivariate analysis) in addition to new tools differing in scope. Both parametric and non-parametric univariate analysis, and supervised and unsupervised multivariate strategies were also used here. Multivariate and univariate designs were additionally used to optimize each variable affecting performance of the different methodologies.

Chapter 8

The work described in this chapter involved using a platform to analyse phospholipids in serum with a view to elucidating their profile in atherosclerotic patients. Phospholipids are a family of lipids constituting the major components of cell membranes and a sizeable portion of the lipid fraction of blood lipoproteins. Most phospholipids contain a diglyceride or monoglyceride unit, a phosphate group and a simple organic molecule such as choline, ethanolamine or serine after which phospholipids classes are named. One exception to this rule is sphingomyelin, which is derived from sphingosine instead of glycerol. The fact that this family of lipids encompasses a large number of metabolites including glycerophosphatidyl-cholines (PC), glycerophosphatidylethanolamines (PE), glycerophosphatidyl-serines (PS) and sphingomyelins (SM), among others, precludes constructing calibration curves for all. For this reason, serum phospholipid profiles were obtained by LC–QTOF MS/MS, a platform that allows identification of phospho-lipids thanks to its high resolution. Phospholipids in

serum can be easily identified from MS/MS information —their fragmentation pattern is quite accurately known. Analyses in the positive and negative ionization modes were used in combination to obtain complementary results.

Sample preparation was intended to isolate phospholipids from serum for preconcentration and removal of potential interferents. This was accomplished by using SPE with cartridges interacting selectively with the phosphate group (zirconia coated silica) in order to selectively retain metabolites containing this functional group. A total of 81 phospholipids were thus detected and identified with this platform; 48 were present in at least 75% of samples (from 140 patients diagnosed with coronary atherosclerosis: 72 with unstable angina/NSTEMI and 68 with stable angina). Statistical analysis of the results allowed four phospholipids present at concentrations significantly differing between patients with stable angina and unstable angina/NSTEMI to be identified, namely: LysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0).

An on-line computer tool called ROCCET was used to build a panel of markers with these four metabolites in order to discriminate between atherosclerotic patients with stable angina and unstable angina/NSTEMI. The panel exhibited a disease prediction ability for the training set of 70.7 and 66.0% for stable angina and unstable angina/NSTEMI, respectively. Also, the area under the ROC curve of the model was 0.715 for the training set and 0.747 for the external validation set, with 66.0% sensitivity and 70.0% specificity. These results warrant a large-scale study to validate the proposed model and testify to the usefulness of the LC–QTOF MS/MS strategy for profiling phospholipids with a view to developing effective aids for clinical diagnosis.

Chapter 9

Taking into account the profile of essential amino acids and their biological relevance, the objective in the study of Chapter 9 was to develop an automated platform for determining essential amino acids in serum and to

identify quantitative differences in amino acids levels between atherosclerotic patients by using the SPE–LC–MS/MS system as instrumental set-up.

The number of metabolites thus identified was large enough to optimize the whole analytical process and to characterize the ensuing method in terms of calibration curve, dynamic range, limit of detection and quantitation, reproducibility and repeatability. The method for determining amino acids was designed with provision for their physico–chemical properties. Thus, because amino acids acquire positive charge in an acid medium, the sample treatment was based on cationic SPE cartridges. Also, their polar nature dictated the use of a HILIC column for chromatographic separation —with an aqueous mobile phase containing a variable proportion of acetonitrile from 90% at the beginning to 40% at the end in addition to 5 mM ammonium formate at pH 6 throughout—, and the fact that the chromatographic phases were incompatible with the solvent needed to elute amino acids from the cartridge (90% acetonitrile, 5% ammonia and 5% water) required elution in the focusing mode. This mode was specially designed for elution with the minimum volume of an eluent other than the mobile phase. The elution step was conducted at a low flow rate in order to minimize its effect on chromatographic separation.

The ensuing optimized automated method was used to analyse samples from 122 atherosclerotic patients (80 with stable angina and 42 who had previously had acute myocardial infarction). Analysis of this batch of samples revealed that ischemia, present in patients with a history of infarction, had a substantial effect on the concentration of two essential amino acids (threonine and lysine), particularly in females. The influence of obesity, an important risk factor for cardiovascular disease, was also examined and obese individuals were found to have increased levels of valine and methionine in serum. This research allowed establishing significant differences in the profile of essential amino acids in serum from individuals diagnosed with atherosclerosis, with special emphasis on the importance of risk factors to improve the discrimination capability.

Chapter 10

This chapter is devoted to the determination of the compounds involved in the tricarboxylic acids (TCAs) cycle in serum by GC–MS after derivatization — a common choice for this family of compounds. The TCA cycle is also known as the "citric acid cycle" or "Krebs' cycle", and takes part in the production of energy from carbohydrates, fats and proteins. Also, it is associated with the production of some amino acids and fatty acids, and is the central metabolic pathway for aerobic organisms.

Sample preparation and determination were optimized for maximal analytical sensitivity and specificity by precipitating proteins with methanol and cleaning up the resulting liquid phase by liquid–liquid extraction with chloroform to remove non-polar interferents, the clean phase then being derivatized by silylation prior to GC–MS analysis. The ensuing method was applied to samples from 223 patients diagnosed with various cardiovascular conditions (172 with significant coronary lesions and 51 with other clinical manifestations but no without coronary lesions). TCA compounds are primary metabolites potentially affected by a number of internal and external factors. For this reason, it is important to consider the influence of cardiovascular risk factors when proposing metabolites of the TCA cycle as potential biomarkers for clinical diagnosis.

This led us to examine the relevance of three risk factors (hypercholesterolemia, obesity and smoking habit) to cardiovascular disease through their influence on the levels of TCA metabolites in serum from patients with coronary lesions. Considering these risk factors proved crucial in order to explain the variability of the target TCA metabolites in patients with coronary lesion as compared to control individuals. In this study, the statistical analysis was based on ROC curves the figures of merit of which were used to assess the discriminating capabilities of TCA metabolites.

Malic acid and α-ketoglutaric acid led to models with high specificity and acceptable sensitivity for discriminating between patients with and without coronary lesions, with provision for obesity —a risk factor that influences the

levels of both metabolites. In any case, the α-ketoglutaric acid model was less robust than the malic acid model.

Chapter 11

The platform for targeted analysis used in this chapter was intended to enable the quantitative determination of peptides of clinical interest. The platform used on-line coupled SPE and tandem mass spectrometry with direct infusion to facilitate the development of fast, highly sensitive and selective methods. Its applicability was assessed by determining the antimicrobial peptide cathelicidin in human serum, using the selective reaction monitoring (SRM) mode with three transitions confirmed by *in silico* fragmentation of the target peptide. The need for an internal standard was avoided by using a dual cartridge configuration to confirm quantitative retention of cathelicidin.

The automated SPE device used, in combination with a polymeric divinylbenzene resin as sorbent, proved efficient in desalting and deproteinizing samples. This is of paramount importance in clinical analysis of biofluids, which frequently involve deproteination with organic solvents. The analytical process took only 12 min per sample. The limits of detection and quantitation were 2.5 and 8.25 μg/L, respectively (0.20 and 0.66 pg on column), and testify to the high efficiency of the SPE–MS/MS system. Repeatability and within-laboratory reproducibility were 2.4% and 2.7%, respectively, and reflected the high precision of the method.

Since cathelicidin takes part in the immune response, where it operates in the first line of defence, the analytical method was validated by application to serum samples from intensive care patients (n = 17) and healthy donors (n = 23). Although the levels of cathelicidin in intensive care patients were higher than those in healthy donors, the differences were not statistically significant.

This approach can be useful to analyze other peptides such as clinical biomarkers and target proteins —following enzymatic hydrolysis to representa-

tive peptides. Thus, with little or no change, it could be applied to any other peptide with considerably reduced costs per analysis as compared with immunoassay. Furthermore, this new approach can be fully automated without the need for a robotic workstation, and also a higher selectivity is obtained because cross-reactivity is absent.

CONCLUSIONES

La investigación que constituye esta Tesis se ha centrado en el desarrollo de aplicaciones basadas en análisis metabolómico global y orientado en el área clínica, destacando la utilidad de la espectrometría de masas para ambas estrategias. Las innovaciones metodológicas constituyen otro de los aspectos que caracterizan la investigación. Las conclusiones más destacadas de la investigación pueden resumirse en los siguientes puntos:

- Las innovaciones desarrolladas para mejorar algunos aspectos débiles del análisis global, que han llevado a concluir que:

    1) Biofluidos poco convencionales como el sudor constituyen una nueva fuente de muestras que debe considerarse en investigación clínica. Las ventajas del sudor en esta área están soportadas en su muestreo no invasivo y en su composición simple. La identificación en sudor de una amplia variedad de familias de metabolitos mediante análisis por LC–QTOF confiere a la exploración de este biofluido un interés muy particular.

    2) El análisis de sudor mediante LC–QTOF tras una simple dilución como única etapa de preparación de la muestra es una opción muy adecuada para obtener una instantánea del metaboloma del sudor, aunque también pueden utilizarse otros protocolos de preparación de la muestra cuando el análisis se oriente a un cierto grupo de metabolitos.

    3) El uso de un sistema automatizado de SPE acoplado a un equipo de LC–QTOF permite mejorar la detección de metabolitos mediante combinación de diferentes sorbentes. Una configuración en serie en la que se acoplan dos sorbentes con mecanismos de retención complementarios ha demostrado ser una alternativa

competitiva para analizar metabolomas complejos mediante una única plataforma, en este caso LC–MS/MS.

4) La aplicación del fraccionamiento en fase gaseosa (GPF) en estudios metabolómicos ha demostrado ser una excelente estrategia para incrementar el número de metabolitos identificados en una muestra dada. De hecho, mientras el modo "auto MS/MS" convencional de adquisición de datos proporcionó información para sólo el 48–57% de las potenciales entidades detectadas, los métodos basados en GPF, mediante la combinación de cuatro rangos para seleccionar iones precursores, proporcionaron información para al menos el 80% de todas las entidades detectadas.

- La aplicación de diferentes plataformas analíticas globales para estudios en nutrimetabolómica o en determinadas enfermedades (aterosclerosis o cáncer de pulmón), han conducido a las siguientes conclusiones:

5) Los perfiles metabolómicos obtenidos mediante una plataforma basada en LC–QTOF utilizando suero de individuos sometidos a diferentes dietas permitió establecer las diferencias metabólicas asociadas a cada dieta. El uso de una etapa de preparación de la muestra para dividir los componentes del suero en sus fracciones polar y no polar resultó una estrategia muy útil para estudiar de forma separada las dos fracciones. Los principales efectos metabólicos de las dietas se adscribieron al metabolismo de los lípidos.

6) La comparación de los perfiles del suero de pacientes con aterosclerosis reveló la presencia de 13 compuestos (bilirrubina y ácidos cólico y desoxicólico, entre otros) cuyos niveles son diferentes dependiendo de que se trate de pacientes con angina estable, infarto agudo de miocardio (AMI) e infarto de miocardio sin elevación del segmento ST (NSTEMI). La bilirrubina ha de-

mostrado ser uno de los metabolitos con mayor influencia para discriminar entre episodios de angina estable y de AMI/NSTEMI. Esta estrategia es también adecuada para la búsqueda de biomarcadores, como lo ha puesto de manifiesto el desarrollo de un panel de compuestos para la predicción del infarto de miocardio con una especificidad del 85.1% y una sensibilidad del 80.8%.

7) En un intento de solventar la principal limitación en el diagnóstico del cáncer de pulmón y conseguir su detección en un estadio temprano, se han propuesto dos paneles de marcadores para la predicción de este tipo de cáncer utilizando sudor. Los dos paneles incluyen maltotriosa y ácido nonanedioico en combinación con γ-GluLeu o MG(22:2). Ambos paneles mejoran de forma significativa la capacidad de metabolitos independientes para la discriminación entre pacientes con cáncer de pulmón y controles. El primer panel se caracteriza por un 100% de especificidad y un 63.6% de sensibilidad. Por tanto, la posibilidad de falsos negativos es del 0%. El segundo panel proporciona valores de sensibilidad y selectividad por encima del 81.0%.

- El desarrollo y aplicación de diferentes plataformas analíticas orientadas ha permitido extraer las siguientes conclusiones:

8) El método desarrollado para el análisis de fosfolípidos permite la identificación de 81 de estos compuestos en suero. La aplicación de esta plataforma para analizar muestras de pacientes ateroscleróticos ha permitido encontrar un grupo de cuatro fosfolípidos —lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0)— que diferencia de forma significativa los pacientes con angina estable de los diagnosticados con NSTEMI. El panel construido con estos cuatro metabolitos tiene una capacidad de predicción en el conjunto de entrenamiento del

70.7 y del 66.0% para pacientes con angina estable y con NSTEMI, respectivamente. Además, la curva ROC obtenida para el modelo presenta un área bajo la curva (AUC) de 0.715 y 0.747 para los conjuntos de entrenamiento y de validación, respectivamente, con un 66.0 y 70.0% de sensibilidad y especificidad, respectivamente.

9) Se ha desarrollado, asimismo, un método para determinar aminoácidos en suero utilizando la plataforma SPE–LC–MS/MS, cuya aplicabilidad se ha demostrado mediante análisis de muestras de pacientes ateroscleróticos diagnosticados con angina estable o infarto agudo de miocardio. Los resultados ponen de manifiesto que la lisina y la treonina se afectan por la isquemia, siendo el efecto más pronunciado en mujeres que en hombres. La influencia de la obesidad, un factor de riesgo clave en enfermedades cardiovasculares, es más significativo para valina y metionina, ambas incrementadas en individuos obesos.

10) Se ha desarrollado un método basado en GC–MS para el análisis de compuestos implicados en el ciclo de los ácidos tricarboxílicos (TCAs), que se ha aplicado a muestras de suero de pacientes con lesiones coronarias y de individuos control. Se ha estudiado la influencia en esta enfermedad de tres factores de riesgo tales como hipercolesterolemia, obesidad y hábito de fumar. La conclusión del estudio es que estos factores deben tenerse en cuenta cuando se propongan metabolitos del ciclo de los TCA como potenciales biomarcadores. El análisis estadístico reveló que los ácidos málico y α-cetoglutárico conducen a modelos de predicción independientes, con buena especificidad y aceptable sensibilidad para la discriminación entre pacientes con y sin enfermedades coronarias.

11) Se ha demostrado la utilidad de una plataforma simple, como la que resulta del acoplamiento SPE–MS/MS, para analizar péptidos mediante su aplicación a la determinación de catelicidina. El péptido se detectó mediante tres transiciones, confirmadas por fragmentación *in silico*. La aplicación del método a pacientes de una unidad de cuidados intensivos y a individuos sanos reveló que no existían diferencias significativas entre ambos grupos, aunque los niveles en los individuos control fueron ligeramente más bajos.

CONCLUSIONS

The research conducted in this Doctoral work was aimed at developing applications based on targeted and untargeted metabolomics analysis in the clinical field, with special emphasis on the usefulness of mass spectrometry for both strategies. Also, the research has introduced some methodological innovations.

The most salient conclusions drawn from this work can be summarized as follows:

- Innovations alleviating the weaknesses of untargeted analysis:

    1) Unconventional biofluids such as sweat can be effective new sources of samples for use in clinical research. The greatest advantages of sweat samples are their simple composition and the fact that they can be obtained non-invasively. Also, this biofluid is potentially useful for identification of metabolites from a wide variety of families by LC–QTOF analysis.

    2) The only sample preparation step required to analyze sweat by LC–QTOF is dilution. This facilitates obtaining a representative snapshot of its metabolome. However, alternative sample preparation protocols can be used to address specific groups of metabolites.

    3) The joint use of automated SPE and LC–QTOF with combinations of different SPE sorbents improves metabolite coverage. A serially arranged combination of two SPE sorbents possessing complementary retention mechanisms seems to be a competitive choice for analyzing complex metabolomes with a single platform.

    4) The use of gas phase fractionation (GPF) in metabolomic studies has proved an effective strategy for increasing the number of

metabolites that can be identified in a given sample. In fact, the conventional "auto MS/MS" mode of data acquisition provided MS/MS information for only 48–57% of the potential entities present, whereas using GPF-based methodology over four different ranges to select precursor ions increased the proportion to at least 80% .

- Use of untargeted analytical platforms in nutrimetabolomics studies and research on specific (atherosclerosis and lung cancer) diseases:

  5) Metabolomic profiling with an LC–QTOF platform of serum from individuals under different diets allowed metabolic differences between diets to be established. Using a sample preparation step to split serum components into a polar and a non-polar fraction proved useful towards examining polar and non-polar compounds separately. The main metabolic effects of intervention diets were found to be associated to lipid metabolism.

  6) A comparison of serum profiles from atherosclerotic patients suffering from different cardiovascular diseases revealed the presence of 13 compounds including bilirubin, cholic acid and deoxycholic acid, the levels of which differed between patients with stable angina, acute myocardial infarction (AMI) and non-ST elevation myocardial infarction (NSTEMI). Bilirubin was among the most useful metabolites to discriminate between stable angina and AMI/NSTEMI events. This strategy is also useful for biomarker identification (e.g., a marker panel allowed myocardial infarction to be predicted with 85.1% specificity and 80.8% sensitivity).

  7) Two marker panels for predicting lung cancer from sweat samples were established that overcome the main shortcoming of lung cancer diagnosis: the difficulty of detection at an early stage. The two panels involve maltotriose and nonanedioic acid in

combination with γ-GluLeu or MG(22:2). Both were found to substantially improve the discrimination capacity of independent metabolites between lung cancer patients and controls. One panel features 100% specificity and 63.6% sensitivity, and hence a 0% false negative rate; the other has specificity and sensitivity above 81%.

- Development and use of different targeted analytical platforms:

  8) The proposed method for phospholipids enabled the identification of 81 phospholipids in serum. Using this platform to analyze samples from atherosclerotic patients allowed the identification of a group of four phospholipids accurately discriminating between patients with stable angina and others diagnosed with NSTEMI, namely: lysoPC(20:5), PC(18:1/18:2), PC(18:0/20:4) and SM(d18:2/14:0). A panel including these four metabolites exhibited a disease prediction capability in the training set of 70.7 and 66.0% for stable angina and NSTEMI patients, respectively. Also, the area under the ROC curve (AUC) for the model was 0.715 for the training set and 0.747 for the validation set, and the resulting sensitivity and specificity 66.0 and 70.0%, respectively.

  9) A method for determining amino acids in serum by SPE–LC–MS/MS was successfully developed and validated on atherosclerotic patients diagnosed with stable angina or acute myocardial infarction. Lysine and threonine levels were found to be affected by ischemic events (particularly in females). Obesity, which is a major risk factor for cardiovascular disease, was especially influential on valine and methionine, which were found at increased levels in obese individuals.

  10) A GC–MS method for determining the compounds involved in the tricarboxylics acid (TCA) cycle in serum was developed and

applied to patients with coronary lesions and control individuals. The role of the risk factors hypercholesterolemia, obesity and smoking habit in the disease was studied, and the influence of the three factors found to be essential with a view to identifying potential biomarkers among metabolites of the TCA cycle. Statistical analysis revealed that malic acid and α-ketoglutaric acid lead to independent prediction models with high specificity and acceptable sensitivity for discriminating between patients with and without coronary lesions.

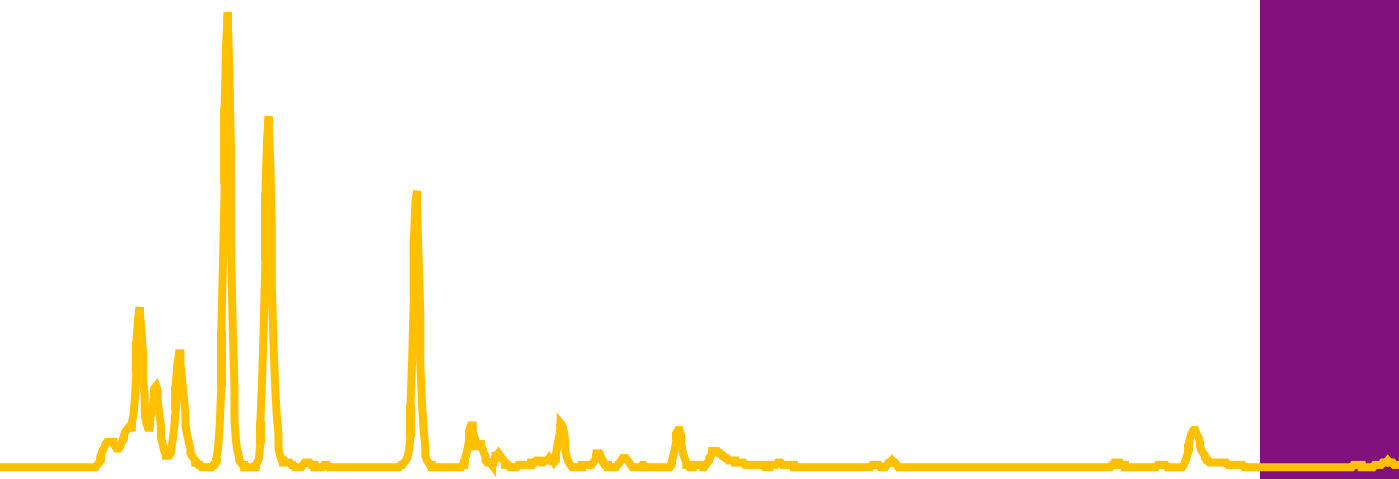
11) The usefulness of SPE–MS/MS for analysing peptides in biofluids was demonstrated by developing an automated platform for assessment and quantitative analysis of cathelicidin. The analyte was detected by using three different transitions that were confirmed by *in silico* peptide fragmentation. Application of this method to intensive care unit patients and healthy individuals revealed the absence of significant differences between the two groups and also that the control individuals had slightly lower concentration levels of the target analytes.

ANNEXES

Anexos

ANNEX I

Other publications co-authored by the PhD student



1. Cholesterol oxidation products in milk: processing formation and determination
M. Calderón-Santiago*, A. Peralbo-Molina, F. Priego-Capote, M.D. Luque de Castro*
European Journal of Lipid Science and Technology 114 (2012) 687–694

---

2. Global metabolomics profiling of human serum from obese individuals by liquid chromatography–time of flight/mass spectrometry to evaluate the intake of breakfast prepared with heated edible oils
C. Ferreiro-Vera, F. Priego-Capote, ***M. Calderón-Santiago****, M.D. Luque de Castro*
Food Chemistry 141 (2013) 1722–1731

---

3. High-resolution mass spectrometry to evaluate the influence of cross-breeding segregating populations on the phenolic profile of virgin olive oils
V. Sánchez de Medina, ***M. Calderón-Santiago****, M. El Riachy, F. Priego-Capote, M.D. Luque de Castro*
Journal of the Science of Food and Agriculture, doi: 10.1002/jsfa.6653

---

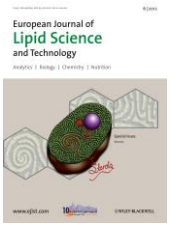
4. Quantitative determination and confirmatory analysis of N-acetylneuraminic and N-glycolylneuraminic acids in serum and urine by solid-phase extraction on-line coupled to liquid chromatography–tandem mass spectrometry
F. Priego-Capote, M.I. Orozco-Solano, **M. Calderón-Santiago**, M.D. Luque de Castro
Journal of Chromatography A, doi: 10.1013/j.chroma.2014.04.051

---

5. Influence of the genotype on the fatty acids composition of virgin olive oils along the ripening process
V. Sánchez de Medina, ***M. Calderón-Santiago****, M. El Riachy, F. Priego-Capote, M.D. Luque de Castro*
Sent to Journal of Agricultural and Food Chemistry

European Journal of Lipid Science and Technology

114 (2012) 687–694



Cholesterol oxidation products in milk: Processing formation and determination

*M. Calderón-Santiago, A. Peralbo-Molina, F. Priego-Capote, M. D. Luque de Castro**

Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, and Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain

Abstract

Cholesterol oxidation products (oxysterols) are commonly present in foods derived from animals. The main limitations for analysis of these products are associated with the long sample preparation protocols that require saponification. In this research, a protocol was optimized for determination of oxysterols in milks under the process to obtain given commercial milk products. UHT milks, powder milks, condensed milk, and evaporated milk were selected for this study. The levels of the different oxysterols are discussed attending to the type of milk. UHT and skimmed powder milks provide the highest concentration of oxysterols. Particularly, high concentrations were found for 7β-hydroxycholesterol (up to 205.6 ng/mL), 5-cholesten-3β-ol-7-one (up to 21.8 ng/mL), cholesterol 5α,6α-epoxide (up to 14.9 ng/mL), and 25-hydroxycholesterol (up to 5.1 ng/mL). An increase in cholesten-3β-ol-7-one and cholesterol 5α,6α-epoxide, resulting from the skimming process, was also observed. Additionally, two different heating modes, microwave-assisted and conventional heating, have been mimicked to study their influence on the

formation of oxysterols in milk. Both heating protocols contributed to the formation of oxysterols but short periods of microwave irradiation (60 s) can lead to levels of oxysterols similar to conventional heating for 5 min.

Practical applications: The results of the research described in this article, supported by previous studies in the field, allow to advice about the strong influence of microwave heating of milk on the formation of toxic oxides from the lipid fraction. Although conventional heating can be more tedious and time consuming, it seems to be healthier.

Food Chemistry

141 (2013) 1722–1731

Global metabolomics profiling of human serum from obese individuals by liquid chromatography–time of flight/mass spectrometry to evaluate the intake of breakfasts prepared with heated edible oils

*C. Ferreiro-Vera, F. Priego-Capote, M. Calderón-Santiago, M. D. Luque de Castro**

Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, and Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain

Abstract

The metabolic profile of human serum after intake of breakfasts prepared with different heated vegetable oils has been studied. Four oils (olive and sunflower oils, pure and enriched with natural and artificial oxidation inhibitors) were subjected to a simulated heated process prior to breakfast preparation. A metabolomics global profiling approach performed on post-basal serum samples revealed statistical differences among individuals based on breakfast intake, and identified compounds responsible for such differences. Serum samples obtained in basal state (control samples) and 2 and 4 h after programmed intakes were analyzed by LC–TOF/MS. The resulting fingerprints were compared and

differences between basal and post-basal states evaluated, observing that the intake of different breakfasts altered the metabolic signature of serum. Analysis models based on PLS algorithms were developed to discriminate individuals in post-basal state for each intervention breakfast. Then, Volcano tests enabled to detect significant molecular entities explaining the variability associated to each breakfast. It is worth emphasizing the importance of fatty acids, their derivatives and phospholipids for tentative identification.

Journal of the Science of Food and Agriculture
doi: 10.1002/jsfa.6653

High-resolution mass spectrometry to evaluate the influence of cross-breeding segregating populations on the phenolic profile of virgin olive oils

V. Sánchez de Medina[1,2,3], M. Calderón-Santiago[1,2,3], M. El Riachy[4], F. Priego-Capote[1,2,3], M. D. Luque de Castro*[1,2,3]*

[1]Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain

[2]University of Córdoba Agroalimentary Excellence Campus, ceiA3, Campus of Rabanales, 14071, Córdoba, Spain

[3]Maimónides Institute for Research in Biomedicine of Córdoba, Reina Sofía University Hospital, University of Córdoba

[4]Lebanese Agricultural Research Institute, Tal Amara, Lebanon

Abstract

BACKGROUND: The growing demand of high quality virgin olive oils (VOOs) has increased the interest in olive breeding programs. Cross-breeding is considered, within these programs, the best strategy to generate new cultivars as an attempt to improve the present cultivars. In this research, the phenolic profile of VOOs from target crosses (Arbequina×Arbosana, Picual×Koroneiki and Sikitita×Arbosana) and their corresponding genitors (Arbequina, Arbosana,

Koroneiki, Picual and Sikitita) has been evaluated using a targeted metabolomics approach.

RESULTS: The phenolic profiles were obtained by LC–QqTOF targeted analysis of thirty-seven phenols or compounds involved in the main pathways for their biosynthesis. Statistical multivariate analysis by Principal Component Analysis (PCA) was applied to study the influence of the genotype on phenols composition. Phenolic compounds with the highest contribution to explain the observed variability associated to genotype were identified through fold change algorithms (cut-off>2.0) and *t*-test analysis.

CONCLUSION: A total of nine phenols (*viz.* quercetin, ligstroside aglycon (*p*-HPEA-EA), demethyloleuropein aglycon, oleuropein aglycon (3,4-DHPEA-EA), hydroxypinoresinol, hydroxytyrosol, phenolic acids such as *p*-coumaric acid, ferulic acid and protocatechuic acid) contributed to explain the observed variability with 99% confidence ($p<0.01$).

Journal of Chromatography A

doi: 10.1013/j.chroma.2014.04.051

Quantitative determination and confirmatory analysis of N-acetylneuraminic and N-glycolylneuraminic acids in serum and urine by solid-phase extraction on-line coupled to liquid chromatography–tandem mass spectrometry

F. Priego-Capote, M. Orozco-Solano, M. Calderón-Santiago, M. D. Luque de Castro*

Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, and Maimónides Institute for Research in Biomedicine of Córdoba, Reina Sofía University Hospital, University of Córdoba, E-14071, Córdoba, Spain

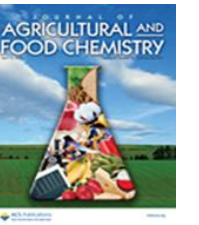
Abstract

N-acetylneuraminic acid (Neu5Ac) and N-acetylglycolylneuraminic acid (Neu5Gc), two acylated derivatives of 9-C carboxylated monosaccharides, are involved in a number of biological processes as modulators of glycoconjugates. A partially automated method is here presented for determination of these sialic acids in the two most important biofluids for clinical analysis: serum and urine. For this purpose, a solid-phase extraction (SPE) workstation was on-line connected to an LC–MS/MS triple quadrupole mass detector. Hydrolysis to release sialic acids bound to glycoconjugates and derivatization were the two steps implemented as sample preparation prior to SPE–LC–MS/MS analysis.

Following thorough optimization of the SPE and LC–MS/MS conditions, the analytical method was validated using the standard addition approach to assess the presence of matrix effects. The proposed method affords detection limits of 0.03 ng/mL and 0.04 ng/mL for Neu5Ac and Neu5Gc, respectively. The precision (expressed as relative standard deviation) was 1.7 and 4.6% for within-day variability, and 4.8 and 7.2% for between-days variability. Accuracy, estimated using spiked (between 1 and 50 ng/mL) and non-spiked samples of both biofluids, ranged from 95.2 to 99.6%. The method was applied to human serum and urine of healthy volunteers, thus showing its suitability for application in both clinical and research laboratories.

Sent to

Journal of Agricultural and Food Chemistry



Influence of the genotype on the fatty acids composition of virgin olive oils along the ripening process

V. Sánchez de Medina[1,2,3], M. Calderón-Santiago[1,2,3], M. El Riachy[4], F. Priego-Capote[1,2,3], M. D. Luque de Castro*[1,2,3]*

[1]Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain

[2]University of Córdoba Agroalimentary Excellence Campus, ceiA3, Campus of Rabanales, 14071, Córdoba, Spain

[3]Maimónides Institute for Research in Biomedicine of Córdoba, Reina Sofía University Hospital, University of Córdoba

[4]Lebanese Agricultural Research Institute, Tal Amara, Lebanon

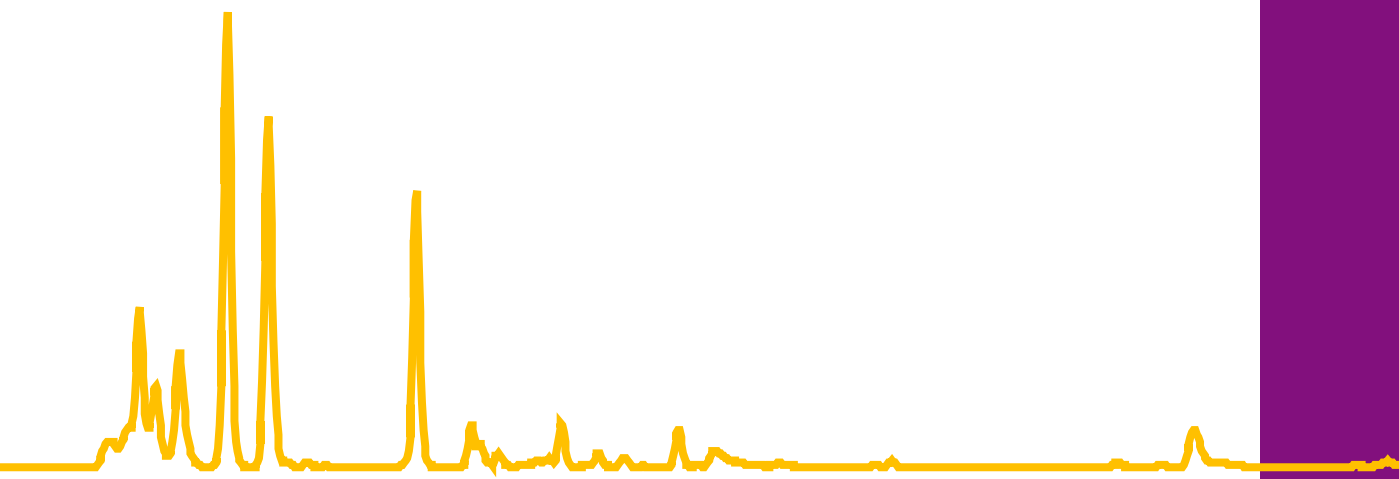
Abstract

The composition of fatty acids (FAs) is one of the most critical aspects that affects the quality of virgin olive oil (VOO), which is related to the balance of concentrations of saturated, monounsaturated and polyunsaturated FAs. The aim of this research was to study the influence of the ripening stage of olive fruits and genotype on the composition of esterified FAs (EFAs) and non esterified FAs (NEFAs) present in VOOs from advanced selections obtained by crosses between Arbequina, Picual and Frantoio cultivars. For this purpose, a method based on gas chromatography with flame ionization detection (GC–FID) was used to

estimate the percent of each type FAs. Statistical unsupervised analysis was carried out by Principal Component Analysis (PCA) to find clustering of samples attending to the ripening stages. Discrimination was observed for VOO samples associated to early and advanced ripening stages according to the concentration of FAs. Statistical analysis by ANOVA test allowed evaluating the contribution of ripening and genotype to explain the variability in the concentration of EFAs and NEFAs with p-value<0.05. Linoleic acid (C18:2) was the FA most influenced by the genotype; practically along the complete ripening process (from yellow or yellowish–green to black color). Furthermore, the highest genetic variability in FAs composition was observed in the ripening stage in which the fruit color is reddish or light violet, since five EFAs (C16:0, C16:1, C18:1, C18:2 and C18:3) were significant to explain this effect. Therefore, the interval that lasts this ripening stage is the most appropriate to compare genotypes in olive breeding programs according to FAs composition.

ANNEX II

Review co-authored by the PhD student



TrAC, Trends in Analytical Chemistry

28 (2009) 1011–1018

The dual trend in histatins research

*M. Calderón-Santiago, M. D. Luque de Castro**

Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain

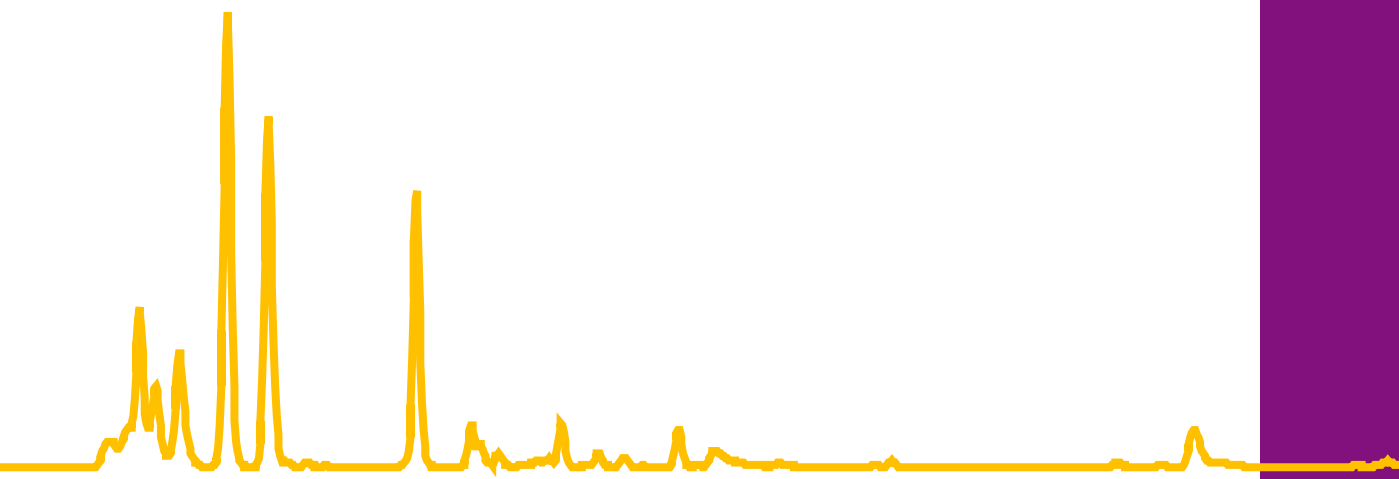
Abstract

We review histatins (Hsts) in order to encourage development of analytical platforms to clarify obscure points in knowledge of this family of antimicrobial and antifungal peptides. To explain the present interest, we outline the number and the nature of Hsts and their known functions (i.e. antimicrobial action, wound closure, biomarkers of stress, satiety, body mass and incipient Alzheimer_s disease, and diagnosis and treatment of addiction, including to cocaine). The two aspects of research on Hsts (i.e. their natural effects on living organisms and their potential use for medical applications, including as biomarkers) make it necessary to develop new analytical methods. The variety of matrices in which Hsts exist (e.g., saliva and tooth-surface-protein pellicle) make it essential to develop new sample-preparation steps and to improve identification and quantitation steps as analytical instrumentation evolves.

In this context, metabolomics studies could be of great interest, as contributions of analytical chemists could be one of the keys to achieving the role that they deserve within "-omics" research.

ANNEX III

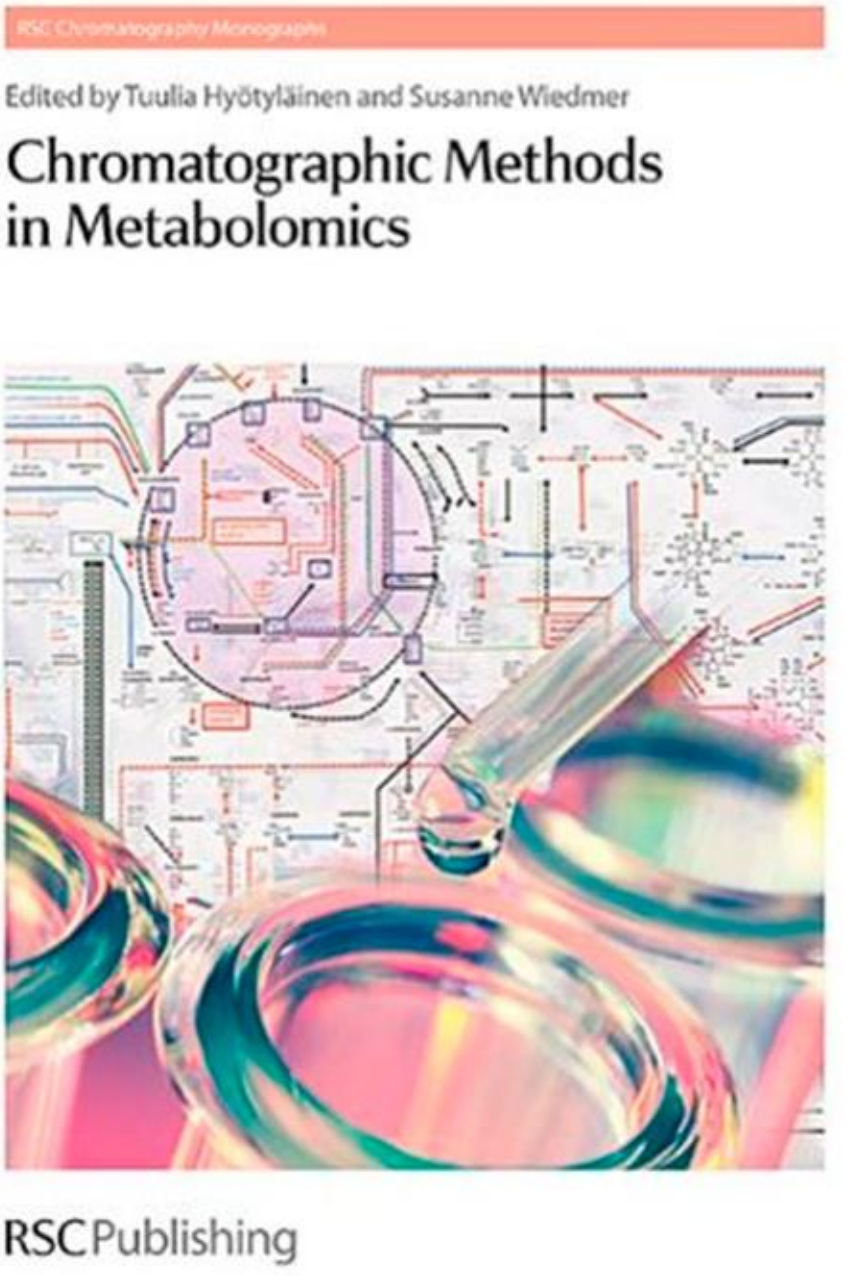
Book chapters co-authored by the PhD student



1. Päivi Pöhö, Maarit Kivilompolo, Mónica Calderón Santiago, Sirkku Jäntti, Susanne K. Wiedmer and Tuulia Hyötyläinen, Chapter 9: Applications, in "Chromatographic methods in metabolomics", edited by Tuulia Hyötyläinen and Susane Wiedner, The Royal Society of Chemistry (RSC) 2013.

---

2. Mónica Calderón Santiago, María Dolores Luque de Castro, Chapter 25: Use of *Lactobacillus spp* to degrade pesticides in milk, in "Processing and impact on active components in food", edited by Victor R. Preedy, Elsevier 2014.


RSC Chromatography Monographs
Edited by Tuulia Hyötyläinen and Susanne Wiedmer
Chromatographic Methods in Metabolomics
RSC Publishing

CHAPTER 9

Applications

PÄIVI PÖHÖ,[a] MAARIT KIVILOMPOLO,[a]
MONICA CALDERON-SANTIAGO,[b] SIRKKU JÄNTTI,[a]
SUSANNE K. WIEDMER[a] AND
TUULIA HYÖTYLÄINEN*[a]

[a] VTT Technical Research Centre of Finland, Tietotie 2, P.O. Box 1000, 02044 VTT, Espoo, Finland; [b] Department of Analytical Chemistry, University of Córdoba Agroalimentary Excellence Campus, Campus of Rabanales, 14071 Córdoba, Spain
*Email: tuulia.tyotylainen@vtt.fi

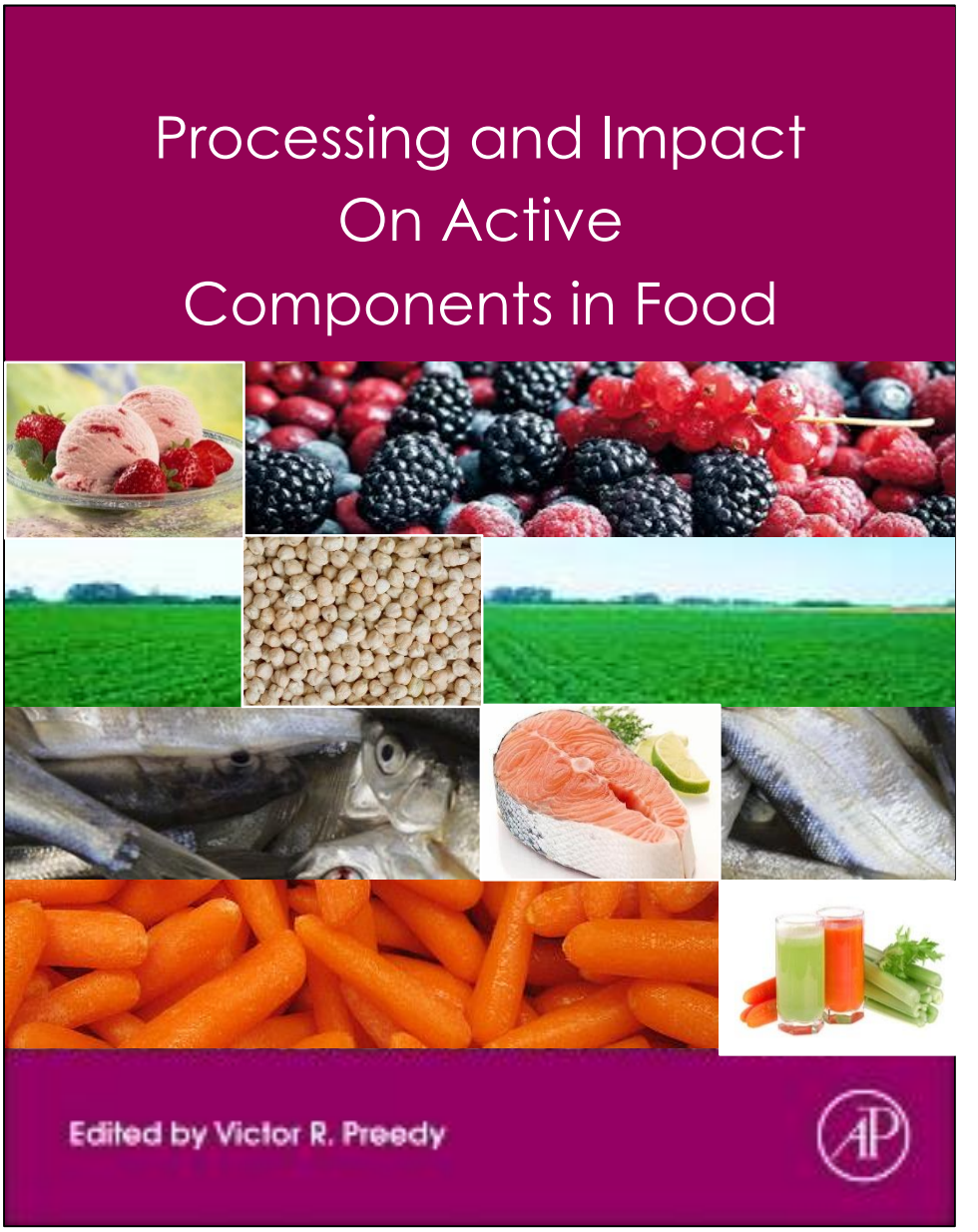
9.1 Introduction

This chapter presents selected applications of metabolomics. The metabolites include compounds from the key metabolic pathways, including lipids, acylcarnitines, acyl-coenzyme A, amino acids, bile acids, carbohydrates, eicosanoids, fatty acids, nucleotides, steroids, sterols and citric acid cycle metabolites. The chapter is divided based on the type of analytes into lipids and polar metabolites, as these groups typically require separate sample preparation methods. Naturally, the type of sample also plays a major role; however, the final analytical methods are typically the same, independent of the type of the sample. The major differences then lie in the sample preparation. Typical illustrative examples are given in the following sections.

The two most common separation techniques used in metabolomics are liquid chromatography (LC) and gas chromatography (GC), while capillary electrophoresis (CE) has been used to a lesser extent. Microfluidic techniques are still under development and have not yet been widely applied for

RSC Chromatography Monographs No. 18
Chromatographic Methods in Metabolomics
Edited by Tuulia Hyötyläinen and Susanne Wiedmer

Published by the Royal Society of Chemistry, www.rsc.org


Processing and Impact
On Active
Components in Food
Edited by Victor R. Preedy
AP

CHAPTER

25

Use of *Lactobacillus spp* to Degrade Pesticides in Milk

Mónica Calderón-Santiago, María Dolores Luque de Castro

Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, Córdoba, Spain

CHAPTER POINTS

- Many foods can be contaminated by pesticides, with milk as the main source of pesticide residues in human diets.
- Microorganisms present in cheese can degrade pesticides during storage.
- Pesticide residues and some nutrients present in milk can be degraded by heating for a long time.
- A strategy to degrade pesticides in milk by use of microorganisms was recently proposed.
- An organophosphorus hydrolase exhibiting high activity on pesticides was obtained by cloning a *Lactobacillus brevis* gen.

INTRODUCTION

The growing demand for agricultural products has raised the need to use pesticides to prevent pest and insect growth. Pesticides are generally deemed safe for crops and animals by virtue of their relatively fast degradation. However, the local environmental conditions can alter their degradation rate and extend their persistence over long periods (Ragnarsdottir, 2000). Although a number of foods can be contaminated by pesticides, those of animal origin are the most difficult to free from these substances; this is especially the case with milk, which is the main source of pesticide residues in human diet owing to the high complexity of its matrix (Kampire *et al.*, 2011). Pesticides can reach milk-producing animals through water, forage, and the environment. Animals can metabolize and detoxify most pesticides except chlorinated hydrocarbons, which are stored in body fat and contaminate meat and milk for human consumption as a result (Li and Bradley, 1969). In fact, pesticides have been found in milk from cows, buffalo, and sheep (Bulut *et al.*, 2011), and even from women, where placenta and milk are the preferential excretory routes of lipophilic pesticides (Siddiqui and Saxena, 1985). Organophosphorus pesticides are the most frequently encountered in milk as a result of their widespread use as herbicides and for cow ectoparasite control; however, they are more rapidly degraded than other pesticides. Figure 25.1 shows the organochlorine and organophosphorus pesticides most frequently found in milk.

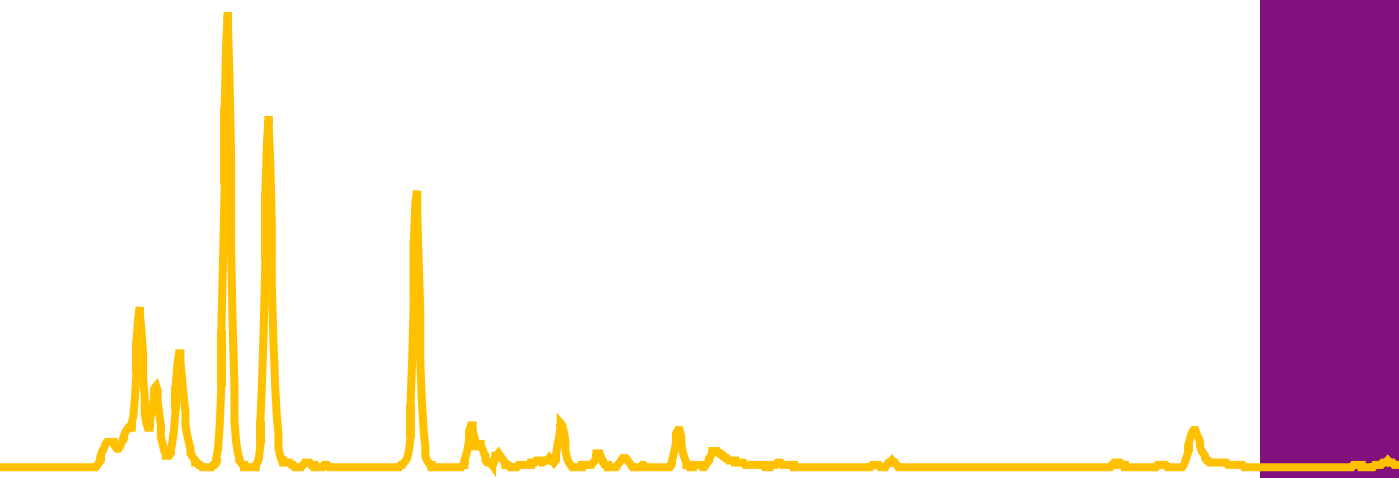
Although pesticides concentrate mainly in the fatty fraction of milk, they rarely disappear when the fat is removed by cooking or other processing steps. All steps by which milk is transformed into dairy products or made suitable for marketing have been studied in depth in order to find ways to efficiently remove pesticides. Also, various strategies for complete pesticide removal from milk have been proposed.

The presence of pesticides in food must be avoided or at least controlled because most of them are mutagenic and teratogenic, and their long-term, low dose exposure is linked to effects on human health such as immunosuppression, hormone disruption, diminished intelligence, reproductive abnormalities, and cancer (Aktar *et al.*, 2009). For these reasons, national and transnational governments have set maximum residue levels of pesticides in or on food and feed of plant and animal origin intended for human or animal consumption. Since pesticide use can not be suppressed altogether without placing food production at risk, a number of strategies have been developed to remove pesticides from contaminated food.

Processing and Impact on Active Components in Food
http://dx.doi.org/10.1016/B978-0-12-404699-3.00025-1

ANNEX IV

Patent co-authored by the PhD student



No Patente: P201331228

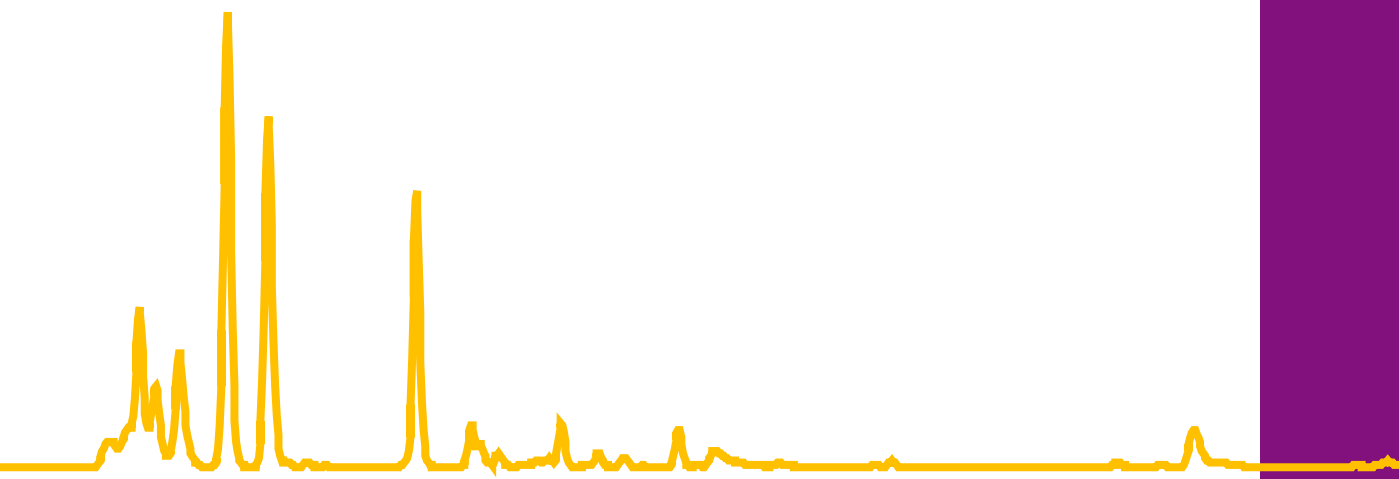
Método de clasificación, diagnóstico y seguimiento de individuos con riesgo de padecer cáncer de pulmón mediante el análisis de sudor

B. Jurado-Gámez, A. Salvatierra-Velázquez, M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro

Los autores de la presente invención han analizado la concentración de los distintos metabolitos del sudor en individuos que no padecen cáncer de pulmón y en individuos que padecen cáncer de pulmón. Han encontrado una serie de marcadores para el diagnóstico de los individuos con cáncer de pulmón, diferenciando los sujetos con cáncer de pulmón de aquéllos que no lo padecen. Esto, entre otras cosas, permitiría hacer un cribado inicial para diferenciar aquellos individuos que serían susceptibles de ser sometidos a otras pruebas diagnósticas, más agresivas o caras, y/o confirmar o apoyar el diagnóstico mediante otras pruebas. Así pues, la presente invención proporciona un método de obtención de datos útiles para la clasificación, diagnóstico y seguimiento de individuos con cáncer de pulmón.

ANNEX V

Communications to congresses and meetings



1. Human sweat metabolomics for lung cancer prediction

M. Calderón-Santiago, F. Priego-Capote, B. Jurado-Gámez, M.D. Luque de Castro

V JORNADAS DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2014

Tipo de evento: Poster en Congreso **Ámbito:** Nacional

2. Study of Coronary Lesions by Analysis of Essential Amino Acids and Citric Acid Cycle in Human Serum

M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M.D. Luque de Castro

XVIII REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA

Úbeda, Jaén, 2013

Tipo de evento: Poster en Congreso **Ámbito:** Nacional

3. Determination of Essential Amino Acids and Tricarboxilic Acid Cycle Metabolites in Human Serum to Study Coronary Lesions

M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M.D. Luque de Castro

XXIII Reunión Nacional de Espectroscopía y VII Congreso Ibérico de Espectroscopía

Universidad de Córdoba, España, 2012

Tipo de evento: Poster en Congreso **Ámbito:** Nacional

4. Metabolomic study of coronary lesions

M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro

III Jornada de Jóvenes Investigadores del IMIBIC

Hospital Universitario Reina Sofia, Córdoba, España, 2012

Tipo de evento: Comunicación oral **Ámbito:** Autonómico

5. Determination of oxysterols in dairy products and study of their formation under heating conditions

M. Calderón-Santiago, A. Peralbo-Molina, F. Priego-Capote, M.D. Luque de Castro

7th Lipidomics Congress (Lipids in all states)

Anglet-Biarritz (Francia), 2010

Tipo de evento: Poster en Congreso **Ámbito:** Internacional

6. Determinación de oxisteroles en derivados lácteos mediante cromatografía de gases con detección por espectrometría de masas y estudio de la influencia de la energía microondas en su formación

M. Calderón-Santiago, A. Peralbo-Molina, F. Priego-Capote, M.D. Luque de Castro

XII Reunión del Grupo Regional Andaluz de la Sociedad Española de Química Analítica

Córdoba, España, 2010

Tipo de evento: Poster en Congreso **Ámbito:** Regional

7. Multiple Reaction Monitoring Verification and Quantitative Analysis of Cathelicidin Biomarker in Human Serum After Automated Sample Cleanup and Deproteinization

F. Priego-Capote, M. Calderón-Santiago, J.M. Mata-Granados, J.M. Quesada-Gómez, M.D. Luque de Castro

58th Annual meeting of American Society of Mass Spectrometry

SALT LAKE CITY, UTAH, USA, 2010

Tipo de evento: Comunicación en congreso **Ámbito:** Internacional

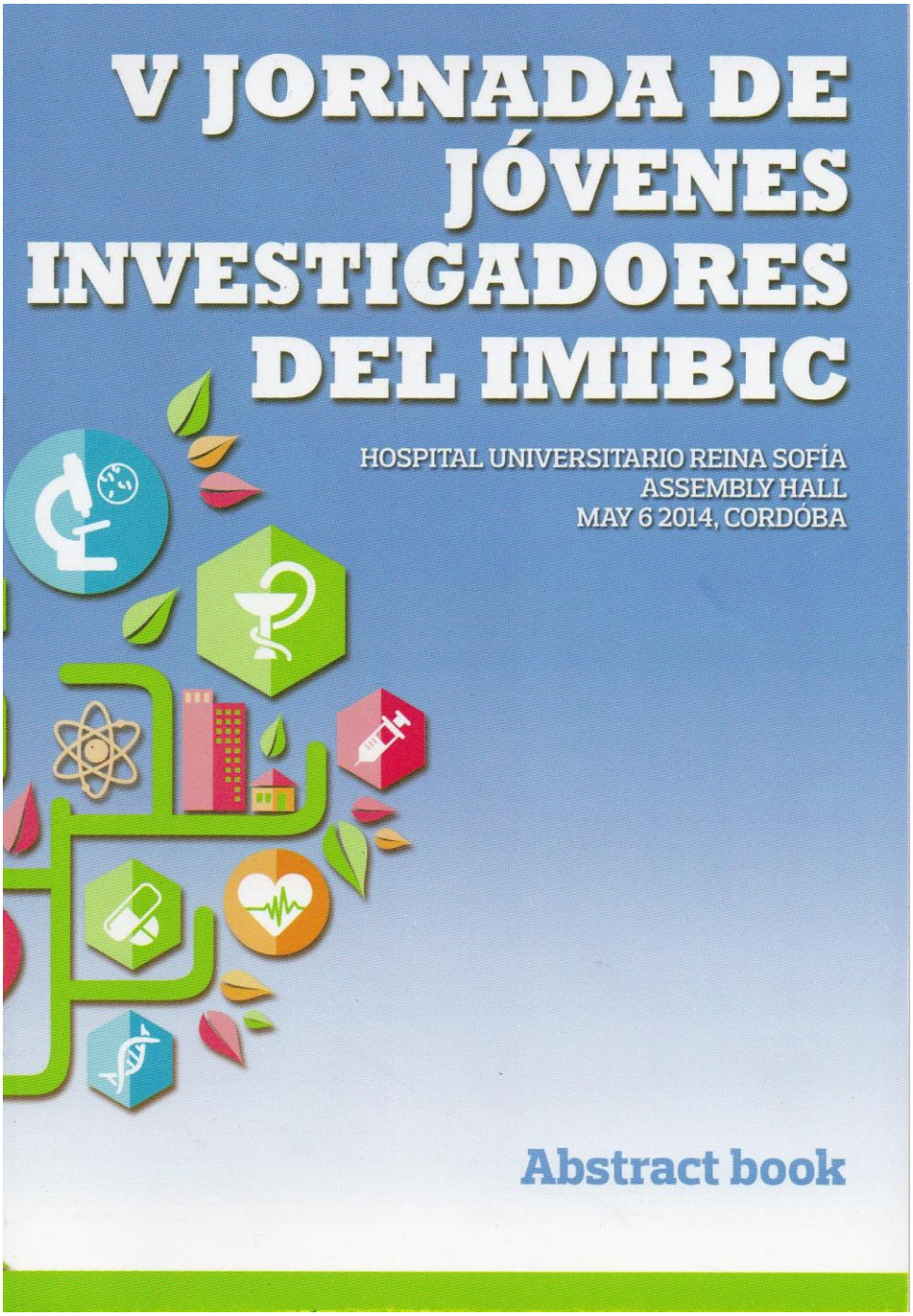
8. Automatic Determination of Cathelicidin in Human Serum by On-Line Solid-Phase Extraction Liquid Chromatography-Triple Quadrupole Mass-Spectrometry with Multiple Reaction Monitoring

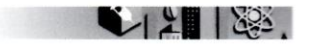
M. Calderón-Santiago, J.M. Mata-Granados, J.M. Quesada Gómez, M.D. Luque de Castro

Flow Analysis XI

Mallorca, Spain, 2009

Tipo de evento: Poster en Congreso **Ámbito:** Internacional


V JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC
HOSPITAL UNIVERSITARIO REINA SOFÍA
ASSEMBLY HALL
MAY 6 2014, CORDÓBA
Abstract book



25 Human sweat metabolomics for lung cancer prediction

Authors: *Mónica Calderón-Santiago, Feliciano Priego-Capote, Bernabé Jurado-Gámez, María D. Luque de Castro*
Group: GC21 *Metabolimics. Identification of bioactive components*

Lung cancer is the carcinogenic disease with the highest mortality rate owing to the advanced stage at which it is usually detected; therefore, biomarkers for lung cancer detection at early stages are needed. On the other hand, sweat has recently gained popularity as a potential tool for diagnostics and biomarker monitoring.

A two-step research has been developed:

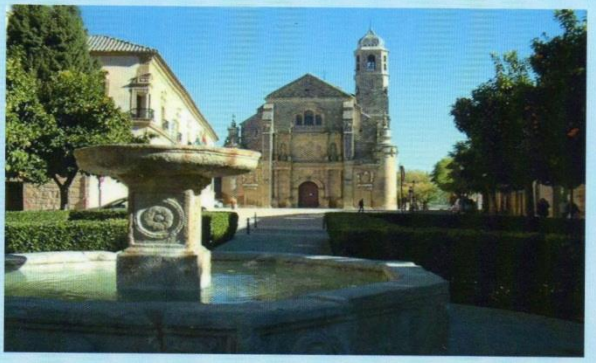
First, an analytical method for analysis of human sweat by liquid chromatography–mass spectrometry (LC–Q-TOF MS/MS) in high resolution mode was developed. Forty one compounds were identified by the MS/MS information obtained with a mass tolerance window below 4 ppm. Amino acids, dicarboxylic acids and other interesting metabolites such as inosine, choline, uric acid and tyramine were identified.

The second step was to use the developed method for analysis of sweat to discriminate between patients with lung cancer versus control individuals (smoker and non-smoker individuals). The capability of the metabolites identified in sweat to discriminate between both groups was studied. Among them, a trisaccharide phosphate presented the best independent performance in terms of the specificity/sensitivity pair (80 and 72.7%, respectively). Additionally, two panels of metabolites were configured using the PanelomiX tool as an attempt to reduce false negatives (at least 80% specificity) and false positives (at least 80% sensitivity). The first panel (100% specificity and 63.6% sensitivity) was composed by nonanedioic acid, γ-GluLeu dipeptide and maltotriose, while the second panel (88.6% specificity and 81.8% sensitivity) included nonanedioic acid, maltotriose and the monoglyceride MG(22:2). As both panels were supported on high sensitivity and specificity values, the proposed approach would be based on the analysis of the four implicated metabolites. The combined use of the two panels would allow reducing the number of cases subjected to confirmatory test with the minimum rate of false negative and positive rates (0% and 18.2%, respectively).

XVIII REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA

SEQA 2013

VI REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE ESPECTROMETRÍA DE MASAS



Úbeda, 16-19 junio 2013

SEQA

S.E.E.M

BAF-01

STUDY OF CORONARY LESIONS BY ANALYSIS OF ESSENTIAL AMINO ACIDS AND CITRIC ACID CYCLE IN HUMAN SERUM

M. Calderón-Santiago[1,2], F. Priego-Capote[1,2], J. G. Galache-Osuna[3], M. D. Luque de Castro[1,2]

[1]Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain
[2]Maimónides Institute of Biomedical Research (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain
[3]Department of Cardiology. Hemodynamic and Interventional Cardiology. University Hospital Miguel Servet. Zaragoza. Spain

Determination of amino acids in serum using a hyphenated SPE–LC–MS/MS system

Amino acids are key regulators of gene expression and the protein phosphorylation cascade, and act as precursors for synthesis of hormones and low-molecular weight nitrogenous substances with enormous biological importance such as nitric oxide, polyamines, glutathione, taurine, etc. The biochemical relevance of amino acids and their levels may be of interest in the study of metabolic and nutritional disorders as well as in other pathologies related with oxidative stress as atherosclerosis. In order to study the influence of atherosclerosis and its risk factors on amino acids levels, a method for determination of leucine/isoleucine, phenylalanine, thryptophan, threonine, lysine, valine and methionine in serum has been developed based on the on-line coupling of an automated solid-phase extraction system (Symbiosis-Pharma system) with an LC–MS/MS device. The method involves an automated sample treatment by solid-phase extraction using anionic cartridges and a focusing elution mode. This elution mode implies elution of the analytes from the cartridge by a small volume of eluent that is incorporated with the LC mobile phase and transports them to the chromatographic column (HILIC) and, then, to the QqQ detector which measures analytes in selected reaction monitoring (SRM) mode. The system is fully automated and the only manual pretreatment of samples is a deproteinization step with methanol.The optimized method was applied to human serum sampled from individuals affected by atherosclerosis (n=122). Eighty of these patients were diagnosed with stable angina, while the rest of the individuals (42) had suffered an acute myocardial infarct. The influence of the presence of atherosclerosis and risk factors such as obesity, hypercholesterolemia and gender on the concentration of essential amino acids was studied by a multifactor analysis of variance, considering interaction between factors. Within the factors with significant influence on amino acids levels, it is noteworthy that the concentrations of lysine (P=0.0335) and threonine (0.0058) were explained significantly by the presence of atherosclerosis, while valine levels (P=0.0167) and methionine levels (P=0.0001) were highly influenced by obesity.

Determination of compounds involved in tricarboxylic acid cycle by GC-MS

Tricarboxylic acid cycle (TCA cycle) is one of the metabolic pathways involved in energy metabolism, producing ATP through the oxidation of acetate from carbohydrates, fats and proteins. Alterations in the TCA cycle have been correlated with numerous pathologies such as cancer, cardiovascular diseases, metabolic syndrome and neurodegenerative disorders, where oxidative stress plays a key role. In order to study the influence of cardiovascular diseases on TCA cycle compounds, a method based on GC–MS for the determination of citric/isocitric, malic, fumaric, succinic, α-ketoglutaric, pyruvic, oxaloacetic and aconitic acids in human serum has been developed.The method involves sample pretreatment, which consists of deproteinization with methanol (1:2 serum–methanol), clean-up with chloroform, evaporation and derivatization with 5:49:1 pyridine–BSTFA–TMCS. Then, 8 µL of the derivatized solution was injected into the GC–MS system in which the target metabolites were determined by single ion monitoring (SIM). The developed method has a detection limit of 10 µg/L and an intra-day and inter-day variability ranging from 3.2–11.4 and 5.8–17.4 %, respectively. In order to study the influence of cardiovascular diseases together with the influence of risk factors on the metabolites, serum samples from 223 patients diagnosed with different cardiovascular problems were analysed. Interactions between categorical (risk factors) and quantitative (TCA cycle metabolites levels) variables were analyzed by a multifactor analysis of variance. Risk factors were very influential on the levels of pyruvic acid, oxaloacetic acid, aconitic acid and fumaric acid while the presence of coronary lesion was found influential only on the levels of aconitic acid, being lower in individuals diagnosed with coronary lesion than in the rest of individuals in the cohort.

XXIII REUNIÓN NACIONAL – VII CONGRESSO IBÉRICO
ESPECTROSCOPÍA
CÓRDOBA 2012 (XXIII RNE – VII CIE)

ACTAS DE LA XXIII RNE – VII CIE

LIBRO DE COMUNICACIONES

UNIVERSIDAD D CÓRDOBA

CÓRDOBA, 17 AL 20 DE SEPTIEMBRE DE 2012

1

Determination of essential amino acids and tricarboxilic acid cycle metabolites in human serum to study coronary lesions

M. Calderón-Santiago[1,2], F. Priego-Capote[1,2], J. G. Galache-Osuna[3], M. D. Luque de Castro[1,2]

[1]Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain
[2]Maimónides Institute of Biomedical Research (IMIBIC), Reina Sofia Hospital, University of Córdoba, E-14071, Córdoba, Spain
[3]Department of Cardiology. Hemodynamic and Interventional Cardiology. University Hospital Miguel Servet. Zaragoza. Spain

Determination of amino acids in serum using a hyphenated SPE–LC–MS/MS system

Amino acids are key regulators of gene expression and the protein phosphorylation cascade, and act as precursors for synthesis of hormones and low-molecular weight nitrogenous substances with enormous biological importance such as nitric oxide, polyamines, glutathione, taurine, etc. The biochemical relevance of amino acids and their levels may be of interest in the study of metabolic and nutritional disorders as well as in other pathologies related with oxidative stress as atherosclerosis. In order to study the influence of atherosclerosis and its risk factors on amino acids levels, a method for determination of leucine/isoleucine, phenylalanine, thryptophan, threonine, lysine, valine and methionine in serum has been developed based on the on-line coupling of an automated solid-phase extraction system (Symbiosis-Pharma system) with an LC–MS/MS device.

The method involves an automated sample treatment by solid-phase extraction using anionic cartridges and a focusing elution mode (*Fig. 1*). This elution mode implies elution of the analytes from the cartridge by a small volume of eluent that is incorporated with the LC mobile phase and transports them to the chromatographic column (HILIC) and, then, to the QqQ detector which measures analytes in selected reaction monitoring (SRM) mode. The system is fully automated and the only manual pretreatment of samples is a deproteinization step with methanol. The method is characterized by a limit of detection ranging from 0.6 to 28.0 ng on column, an SPE recovery from 96.9 to 99.9% and repeatability/reproducibility from 2.3–6.0 and 3.9–10.4 %, respectively, both expressed as standard deviation. The optimized method was applied to human serum sampled from individuals affected by atherosclerosis (n=122). Eighty of these patients were diagnosed with stable angina, while the rest of the individuals (42) had suffered an acute myocardial infarct. The influence of the presence of atherosclerosis and risk factors such as obesity, hypercholesterolemia and gender on the concentration of essential amino acids was studied by a multifactor analysis of variance, considering interaction between factors. Within the factors with significant influence on amino acids levels, it is noteworthy that the concentrations of lysine (P=0.0335) and threonine (0.0058) were explained significantly by the presence of atherosclerosis, while valine levels (P=0.0167) and methionine levels (P=0.0001) were highly influenced by obesity.

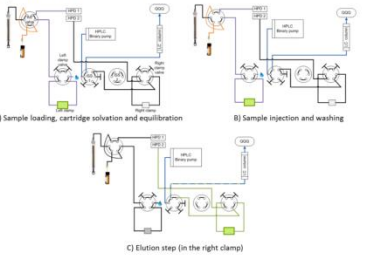


Figura 2. Instrumental configuration including the main steps involved in the SPE step.

Conclusions

The ischemia phenomena, present in patients who had suffered an infarct, has revealed a significant influence on the concentration of two essential amino acids such as threonine and lysine. This effect was more pronounced in female than in male individuals. The influence of obesity, as an important risk factor in cardiovascular diseases, has also been elucidated indicating higher levels of valine and methionine in serum from obese individuals.

Determination of compounds involved in tricarboxylic acid cycle by GC-MS

Tricarboxylic acid cycle (TCA cycle) is one of the metabolic pathways involved in energy metabolism, producing ATP through the oxidation of acetate from carbohydrates, fats and proteins. Alterations in the TCA cycle have been correlated with numerous pathologies such as cancer,

cardiovascular diseases, metabolic syndrome and neurodegenerative disorders, where oxidative stress plays a key role. In order to study the influence of cardiovascular diseases on TCA cycle compounds, a method based on GC–MS for the determination of citric/isocitric, malic, fumaric, succinic, α-ketoglutaric, pyruvic, oxaloacetic and aconitic acids in human serum has been developed. The method involves sample pretreatment, which consists of deproteinization with methanol (1:2 serum–methanol), clean-up with chloroform, evaporation and derivatization with 5:49:1 pyridine–BSTFA–TMCS. Then, 8 µL of the derivatized solution was injected into the GC–MS system in which the target metabolites were determined by single ion monitoring (SIM). The developed method has a detection limit of 10 µg/L and an intra-day and inter-day variability ranging from 3.2–11.4 and 5.8–17.4 %, respectively. In order to study the influence of cardiovascular diseases together with the influence of risk factors on the metabolites, serum samples from 223 patients diagnosed with different cardiovascular problems were analysed. Interactions between categorical (risk factors) and quantitative (TCA cycle metabolites levels) variables were analyzed by a multifactor analysis of variance. Risk factors were very influential on the levels of pyruvic acid, oxaloacetic acid, aconitic acid and fumaric acid while the presence of coronary lesion was found influential only on the levels of aconitic acid, being lower in individuals diagnosed with coronary lesion than in the rest of individuals in the cohort. The analysis of two-factor interactions revealed the importance of certain risk factors to explain serum levels of the target TCA cycle metabolites. Two of the main interactions were coronary lesion × smoking habit and coronary lesion × obesity, with highly-significant influence on pyruvic acid ($p = 0.0029$) as well as malic acid ($p = 0.0214$) and α-ketoglutaric acid ($p = 0.0021$). These results were validated by ROC curves and by estimation of the pAUC, recommended when the performance test only takes place in high specificity or high sensitivity regions. According to the ROC curves (*Fig. 2*), aconitic and pyruvic acids were able to discriminate patients with or without coronary lesions (with or without smoking habit for pyruvic acid) with high specificity (92 and 91%, respectively), but with low sensitivity (21 and 17%, respectively). The situation was different for the other two TCA cycle metabolites whose concentration profiles were explained significantly by the interaction between coronary lesion and obesity. Thus, malic acid was able to discriminate individuals with coronary lesions with specificity of 92% and sensitivity of 49%, while α-ketoglutaric acid reported values of 93% specificity and 56% sensitivity. The α-ketoglutaric acid seems to have a higher discrimination capability than malic acid. However, these results need to be supported on the pAUC parameter of the two ROC curves, which was 2.3% for α-ketoglutaric acid and 4.2% for malic acid. These values enable to conclude that α-ketoglutaric acid discrimination capability is less robust than that observed for malic acid.

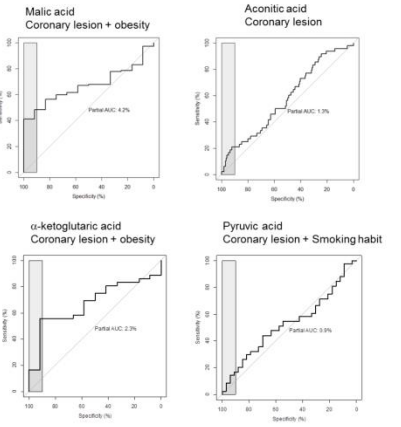


Figure 3. ROC curves of malic, aconitic, α-ketoglutaric and pyruvic acids for the one- factor or two- factors combination for which each acid was significant. The vertical grey box corresponds to the pAUC region (between 90 and 100% specificity).

Conclusions

The relevance on cardiovascular diseases of three risk factors such as hypercholesterolemia, obesity and smoking habit has been studied through their influence in serum levels of TCA cycle metabolites for patients diagnosed with coronary lesions. The consideration of these risk factors was crucial to explain the variability of the target TCA cycle metabolites for patients with coronary lesion versus control individuals. Both malic acid and α-ketoglutaric acid were able to discriminate between patients with and without coronary lesions considering the obesity with a sensitivity of 49 and 56 % and a specificity of 92–93 %, respectively.


III Jornada de Jóvenes Investigadores del IMIBIC
Salón de actos del Rectorado de la UCO
Avda Medina Azahara, 7
CÓRDOBA 16 ABRIL 2012
+info: www.imibic.org

IMIBIC

05 METABOLOMIC STUDY OF CORONARY LESIONS

Autor/res: Calderón-Santiago M, Priego-CapoteF, Luque de Castro MD.
Grupo IMIBIC: Metabolómica. Identificación de componentes bioactivos.
Comunicación oral

Nowadays, metabolomics is one of the most used 'omics' to study diseases. We have used two metabolomics strategies in order to study metabolic changes in serum from individuals diagnosed with coronary lesions (CLs) and the influence of risk factors such as smoking habit, obesity, hypercholesterolemia, and hypertension on them: targeted analysis, which involves qualitative and quantitative analysis of a small group of metabolites, and global metabolomic profiling to obtain a comprehensive profile of the metabolome. On one hand, essential amino acids levels have been determined by SPE-HPLC-MS/MS to establish differences in CL patients diagnosed with stable angina or acute myocardial infarction, as the occurrence of thrombi affected the concentration of target metabolites. The influence of both obesity —as an important risk factor in cardiovascular diseases—, gender and hypercholesterolemia–obesity has also been elucidated. No significant influence was observed of hypertension or smoking habit on variability of essential amino acids levels.
Furthermore, global metabolomic profiling of samples from CL patients diagnosed with stable angina, non-ST segment elevation coronary syndrome and acute myocardial infarction has led to elucidate, (i) metabolites with different behaviour in each case, which could enable discrimination between the three diagnostics, and (ii) information about metabolic changes related with CLs occurrence. Profiles from the target serum samples obtained by LC–TOF/MS have allowed detection of 31 and 51 significant molecular entities (the most relevant molecules were bilirubin, biliverdin, uric acid, glutamine, lysophosphocholine and docosahexaenoic acid) in negative and positive mode, respectively. These entities were used to build a partial least squares discriminant analysis (PLS-DA) model for discrimination between patients with stable angina and those with non-ST elevation coronary syndrome (NSTECS) and acute myocardial infarction (AMI); thus obtaining accuracy of 81% in discrimination between AMI and stable angina and accuracy of 79% in discrimination between NSTECS and stable angina.

Keywords: Metabolomic, Atherosclerosis, Amino acids, Acute Myocardial Infarction, Mass spectrometry

Groupe d'Etude et de Recherche en Lipidomique

7TH LIPIDOMICS CONGRESS
« LIPIDS IN ALL STATES »

3rd-6th October 2010 Anglet-Biarritz, France

P8 - Determination of oxysterols in dairy products and study of their formation under heating conditions

Mónica Calderón-Santiago[1], Ángela Peralbo-Molina[1], Feliciano Priego-Capote[1] and M. D. Luque de Castro[1]

[1]Department *of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*

Email: b42casam@uco.es

Cholesterol, because of its chemical structure, is a compound sensitive to oxidation, generating a series of derived products that are collectively termed oxysterols. Oxysterols can be formed in processing of foods under different physico-chemical conditions (heating, cooling, etc.). The clinical interest aims at oxysterols can be absorbed in the small intestine and incorporated into chylomicrons. Additionally, oxysterols have shown different toxic effects, some of which are alteration of the structure and function of the cellular membranes or changes in the activity and expression of enzymes involved in cholesterol biosynthesis [1]. All these characteristics of oxysterols make monitoring the presence of them in foods a necessity from the standpoint of food safety. Furthermore, the oxidation mechanism is not well known and must be clarify [2].

Oxysterols determination in food products is a challenge in analytical chemistry due to the low concentration of these compounds in processed foods and the complexity of the lipid fraction. In this research, a method for determination of oxysterols in milk-derived products was developed. Additionally, a study about the manufacture process on the formation of oxysterols was carried out. This study was based on elucidating the influence of microwave energy *versus* conventional heating (both used in manufacture processes) on their formation: the higher influence of the former on degradation leads to alert about the importance of microwave energy (time and power) on the quality of the dairy products of daily intake.
After mimicking the processing of products, treated samples were prepared by liquid–liquid extraction following saponification in order to isolate oxysterols present in the unsaponifiable fraction. This fraction was cleaned by solid-phase extraction with aminopropyl cartridges in order to remove interferences and preconcentrate oxysterols. Finally, oxysterols were derivatized to trimethylsilyl ethers for subsequent analysis by CG-MS in single ion monitoring (SIM) mode for selective determination of the oxysterols panel.

References:

1. Valenzuela, A., Sanhueza, J., Nieto, S. (2002) Cholesterol oxides (oxisterols): factors conditioning their formation, biological effects and content in foods, *Rev. Chil. Nutr. 29*, 116.
2. Smith, L. (1996) Review of progress in sterol oxidation, Lipids *31*, 453–488

REUNIÓN DEL GRUPO REGIONAL ANDALUZ DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA

XII GRASEQA 2010

Córdoba, 10 y 11 de junio

LIBRO DE RESÚMENES

UNIVERSIDAD D CÓRDOBA

GRASEQA 2010

SC-37

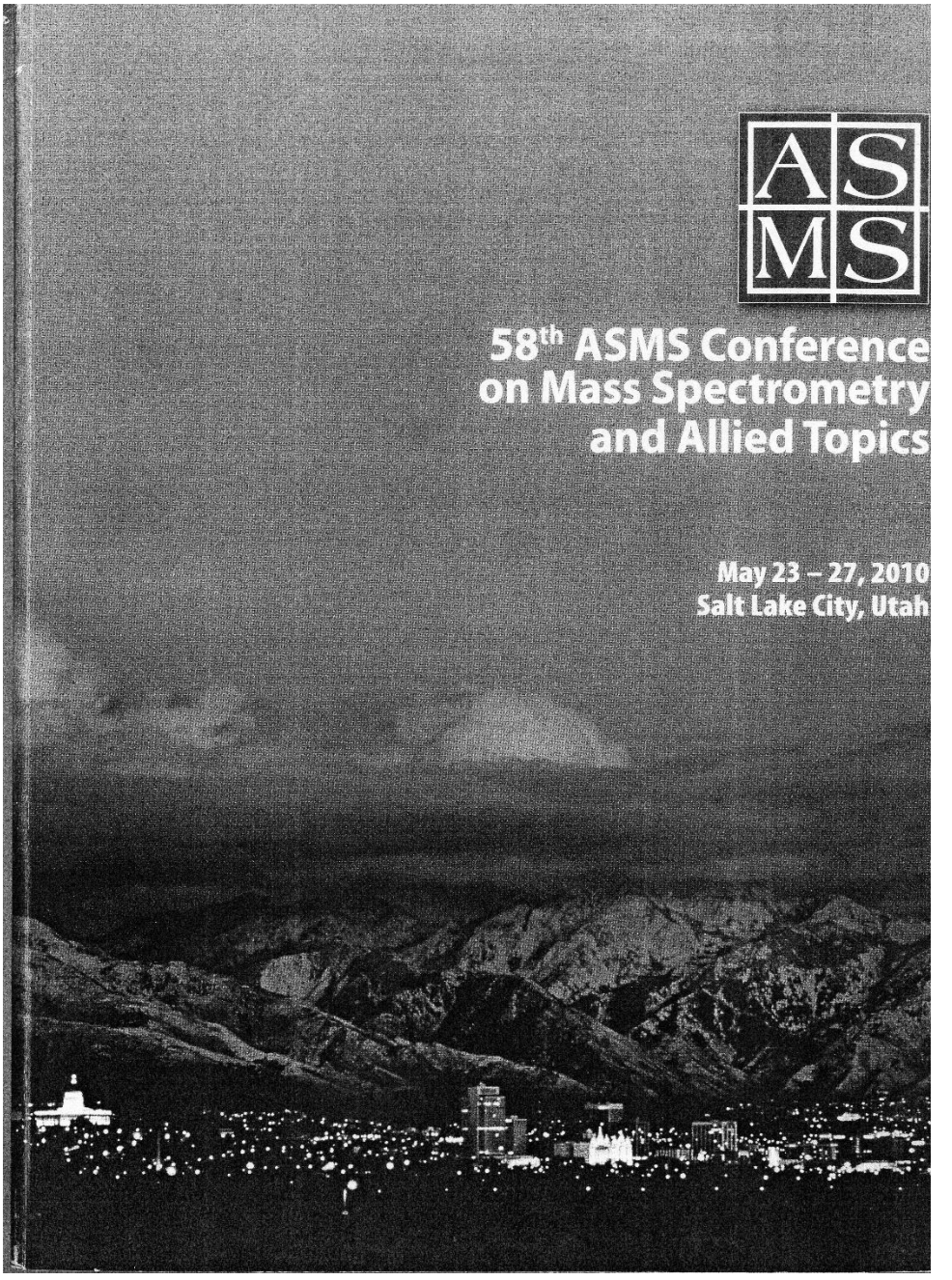
Determinación de oxisteroles en derivados lácteos mediante cromatografía de gases con detección por espectrometría de masas y estudio de la influencia de la energía microondas en su formación

M. Calderón-Santiago[a,b], A. Peralbo-Molina[a,b], F. Priego-Capote*[a,b], M.D. Luque de Castro[a,b]

[a]*Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, E-14071, Córdoba, Spain*
[b]*Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba, E-14071, Córdoba, Spain*

Los procesos de oxidación en la industria alimentaria son indicadores nutricionales y de calidad. Los oxiesteroles o productos derivados de la oxidación del colesterol constituyen un ejemplo claro de este tipo de procesos, ya que exhiben actividades biológicas indeseables tales como citotoxicidad o la aparición de mutaciones. Los oxiesteroles han demostrado ser biológicamente activos en experimentos *in vitro* actuando como ligandos de los receptores X del hígado (*liver X receptors,* LXRs) que regulan la síntesis enzimática de colesterol en la ruta SREBP-2 y aceleran la degradación de la enzima HMG-CoA reductasa. A pesar de esta notable actividad *in vitro*, su papel *in vivo* es todavía objeto de debate. Teniendo en cuenta la presencia de colesterol en productos alimentarios de origen animal tales como leche, carne o huevos, la formación de oxiesteroles puede activarse mediante el procesado de alimentos a altas temperaturas o su almacenamiento prolongado.

La determinación de oxiesteroles en productos alimenticios supone un reto en química analítica debido a la baja concentración a la que se encuentran y a la presencia de interferencias procedentes de la fracción lipídica. En este caso se realizó un estudio de determinación de oxisteroles en distintos derivados lácteos y de la influencia de la energía de microondas en su formación. Para ello, se llevó a cabo un proceso de simulación de la formación de estos compuestos en un hipotético procesado del alimento. El proceso analítico se inició con la extracción líquido–líquido de la fase lipídica seguido de su saponificación con el fin de aislar los oxiesteroles contenidos en la fracción insaponificable. Esta fracción se trató mediante extracción en fase sólida utilizando cartuchos empaquetados con aminopropilo con el fin de eliminar interferencias y preconcentrar los oxiesteroles. Finalmente, se llevó a cabo la derivatización de los oxisteroles mediante sililación para su posterior análisis mediante CG-MS en modo SIM ("Single Ion Monitoring") con el fin de determinar de forma selectiva el perfil de los siete oxiesteroles seleccionados para este estudio.


ASMS
58th ASMS Conference
on Mass Spectrometry
and Allied Topics
May 23 – 27, 2010
Salt Lake City, Utah

Preliminary data:

A multivariate study was developed in order to operate under optimum conditions in the SPE step (testing different sorbent materials) and mass spectrometry detection. As a result of this study, the approach enables completion of the analysis in approximately 12 minutes that means a sampling frequency of 5 analyses/hour. The accuracy of the method and potential matrix effects were assessed by analysis of non-spiked serum samples in order to estimate the recovery of cathelicidin with a two-cartridge configuration of the Prospekt-2 system. Two cartridges were located in serial, so the amount of analyte no retained in the first cartridge was retained in the second. Recovery in this system was calculated resulting in 80% retention. As quantitative retention in the cartridge was ensured, determination of cathelicidin was internally validated. In this way, synthetic peptides labeled with stable isotopes as internal standards are not mandatory to validate quantitation. Detection and quantitation limits were 2.5 ng/L and 8.25 ng/L, respectively (0.2 pg and 0.66 pg on column), while repeatability and within laboratory reproducibility were 2.4 and 2.7%, respectively. Serum samples isolated from intensive care patients (n=17) and healthy blood donors as control group (n=23) were collected in order to compare cathelicidin levels and establish normal concentrations in human serum. The mean age of population was 41±16 year-old formed by 26 men and 14 women. The group of intensive care patients possessed a higher variability in cathelicidin profile as compared to the group of control individuals, which can be represented by a normal distribution. In addition, the average cathelicidin level was higher in intensive care patients (48.4 ± 29.8 ng/mL) *versus* that found in the control group (34.9 ± 18.7 ng/mL). However, no statistical differences were found running t-test (p=0.22), taking into account standard deviation values.

Novel aspects: An approach for verification and quantitative analysis of cathelicidin, a potential biomarker, based on targeting mass spectrometry.


FLOW ANALYSIS XI
September 14-18, 2009
Pollensa. Mallorca. Spain
PC

P046

AUTOMATIC DETERMINATION OF CATHELICIDIN IN HUMAN SERUM BY ON-LINE SOLID-PHASE EXTRACTION LIQUID CHROMATOGRAPHY–TRIPLE QUADRUPOLE MASS-SPECTROMETRY WITH MULTIPLE REACTION MONITORING

M. Calderón-Santiago[a], J.M. Mata-Granados[a,b], J.M. Quesada Gómez[b,c], M.D. Luque de Castro[a]

[a] *Department of Analytical Chemistry, Annex C-3, Campus of Rabanales, University of Córdoba, E- 14071, Córdoba, Spain*

[b] *Department of I+D+i, Sanyres group,E-14012, Córdoba, Spain*

[c] *Mineral Metabolism Unit, Hospital Reina Sofia of Córdoba, E-14004, Córdoba, Spain*

Cathelicidin, also known as LL-37, is one of the human antimicrobial peptides that take part in the innate immune system, operating in the first line of defense. LL-37 is a peptide with 37 amino acids and with α-helix secondary structure. This antimicrobial peptide acts also as chemiatractant of monocytes, neutrophiles and T cellules, and stimulates endothelial proliferation. It is synthesized in macrophages, neutrophiles and epithelial cellules. The cathelicidin gen requires the presence of the active form of vitamin D (1,25 $(OH)_2D_3$) to be activated. This fact makes that people with low levels of vitamin D may present low levels of cathelicidin, and a treatment with vitamin D can increase levels of cathelicidin. It can be of interest to know cathelicidin levels in patients with some disease, mainly unit-care patients, and also to establish normal levels of this peptide in healthy population.

With the aim of providing an appropriate tool for these studies, a fast automated method based on hyphenated solid-phase extraction (SPE) and liquid chromatography–tandem mass spectrometry has been developed and proved to be highly selective and sensitive to determine cathelicidin in serum.

The SPE step was developed using a Prospekt 2 system and Hysphere Resin GP as sorbent, thus providing cleanup/preconcentration of the target analyte in less than 10 minutes. 1 µl of 10% formic acid solution was added to 100 µl of human serum, and then this mixture was diluted ten times with water. 200 µl of this solution was directly

injected in a Prospekt system. Methanol and 0.1% formic acid were used as solvation and load solvents, respectively. Cathelicidin was retained in the cartridge and then, the initial chromatographic mobile phase (70:30 50% methanol–100% methanol, both with 0.2% acetic acid and 0,1% formic acid as ionizing agents) eluted the peptide to the analytical column (synergy hydro RP 100x2mm, 2.5 µm particle size) for chromatographic separation in 5 minutes. Linear gradients were programmed to obtain 100% methanol at 5 minutes. The column temperature was 40 °C and the total analysis time 15 minutes. The recovery was 96.8%.

Automation of sample preparation by using on-line SPE with automatic valve switching and cartridge exchange has proved an excellent approach for unattended analysis of the target compounds in human serum samples. Minimising human intervention makes the method easy to apply and improves reproducibility and accuracy. In addition to full automation, the proposed approach provides a closed system for maximum protection against degradation and high precision because the number of sample transfer steps is significantly reduced.

Abbreviations

ACE, automated cartridge exchange

ACN, acetonitrile

ACS, acute coronary syndrome

AMI, acute myocardial infarction

ANOVA, analysis of variance

AQUA, absolute quantitation

AUC, area under the curve

BCAAs, branched chain amino acids

BMI, body mass index

BPC, base peak chromatogram

BSTFA, bis-(trimethylsilyl)-fluoroacetamide

BTP, black tea polyphenols

CAD, coronary arthery disease

CD, Crohn's disease

CE, capillary electrophoresis

CI, confidence interval

CoA, coenzyme A

COX-2, cyclooxygenase 2

CPK, creatinin phosphokinase

CVD, cardiovascular disease

CYP, cytochrome P450

DG, diglyceride

EBC, exhaled breath condensate

ECG, electrocardiogram

EI, electron impact ionization

EIC, extracted ion chromatogram

ELISA, enzyme-linked immunosorbent assay

EPA, eicosapentaenoic acid

ESI, electrospray ionization

FID, flame ionization detector

FT-IR, Fourier transformation infrared

FWHM, full width at half maximum

GC, gas chromatography

GPF, gas phase fractionation

HDL, high density lipoprotein

HETE, hydroxyeicosatetraenoic acid

H-FABP, human fatty acid binding protein

HILIC, hydrophilic interaction liquid chromatography

HLB, hydrophilic–lipophilic balance

HMDB, Human Metabolome Database

HMUFA, high fat content diet with monounsaturated fatty acids

HO, hemeoxygenase

HODE, hydroxyoctadecadienoic acid

HP-921, hexakis (1H, 1H, 3H-tetrafluoropropoxy) phosphazine

HPD, high pressure dispenser

HpODE, Hydroperoxyoctadecadienoic acid

HSFA, high fat content diet with saturated fatty acids

Ile, isoleucine

IS, internal standard

IT, ion trap

KEGG, Kyoto Encyclopedia of Genes and Genomes

LC, liquid chromatography

LDL, low density lipoprotein

Leu, leucine

LFHCC, low fat content diet supplemented with oleic acid

LFHCCn-3, low fat content diet supplemented with omega 3

LIPGENE, Diet, genomics and the metabolic syndrome: An integrated nutrition, agro-food, social and economic analysis

LLE, liquid-liquid extraction

LOD, limit of detection

LOQ, limit of quantitation

LPC, lysophosphatidylcholine

LPE, lysophosphatidylethanolamine

Lys, lysine

MALDI, matrix-assisted laser desorption ionization

MANOVA, multivariate analysis of variance

Met, methionine

METLIN, Metabolites and Tandem MS Database

MetS, metabolic syndrome

MF, molecular feature

MG, monoacylglycerol

MIDAS, autosampler

MM, mix mode

MPP, mass profiler professional

MS, mass spectrometry

MUFA, monounsaturated fatty acid

NMR, nuclear magnetic resonance

NSCLC, non-small cell lung cancer

NSTEACS, non-ST elevation acute coronary syndrome

NSTEMI, non-ST elevation myocardial infarction

NU-AGE, new dietary strategies addressing the specific needs of elderly population for a healthy aging in Europe

OGTT, oral glucose tolerance test

PA, glycerophosphatidic acid

pAUC, partial area under the curve

PC, glycerophosphatidylcholine

PCA, principal component analysis

PE, glycerophosphatidylethanolamine

PG, glycerophosphatidylglyceride

PGE3, prostaglandin E3

Phe, phenylalanine

PLS, partial least squares

PLS-DA, partial least squares discriminant analysis

PLTP, plasma phospholipid transfer protein

PM, personalized medicine

PS, glycerophosphatidylserine

PUFA, polyunsaturated fatty acid

QC, quality control

QqQ, triple quadrupole

QTOF, quadrupole–time of flight

ROC, receiver-operating characteristic

ROS, reactive oxygen species

RSD, relative standard deviation

RT, retention time

SA, stable angina

SD, standard deviation

SFA, saturated fatty acid

SH, strong hydrophobic

SM, sphingomyelin

SPE, solid-phase extraction

SPME, solid-phase microextraction

SIM, selected ion monitoring

SRM, selected reaction monitoring

T2D, type 2 diabetes

T2DM, type 2 diabetes mellitus

TCA, tricarboxylic acid

TG, triglyceride

Thr, threonine

TIC, total ion current

TLC, thin-layer chromatography

TMCS, trimethylchlorosilane

Trp, tryptophan

UC, ulcerative colitis

UHD, ultrahigh definition

UPLC, ultra performance liquid chromatography

Val, valine

VLDL, very low density lipoprotein

VOC, volatile organic compounds

IMIBIC
INSTITUTO MAIMÓNIDES DE INVESTIGACIÓN BIOMÉDICA DE CÓRDOBA

UNIVERSIDAD DE CÓRDOBA

ceiA3
campus de excelencia internacional agroalimentario