UNIVERSIDAD SAN PABLO CEU

FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA
SECCIÓN DE QUÍMICA ANALÍTICA

EXPLORING THE CAPABILITIES IN THE DIVERSE ANALYTICAL PLATFORMS IN METABOLOMICS.

EXPLORANDO LAS CAPACIDADES DE DISTINTAS HERRAMIENTAS ANALÍTICAS EN ESTUDIOS DE METABOLÓMICA.

Doctoral dissertation

Presented by:

ALMA CRISTINA VILLASEÑOR SOLIS

Supervisor:

PROF. CORAL BARBAS

MADRID, 2014.

AGRADECIMIENTOS.

Quisiera dedicar y agradecer a Dios por permitirme concluir este proyecto que ha sido uno de los más importantes en mi vida.

Este trabajo ha sido el resultado de cinco años de intenso esfuerzo y dedicación, durante los cuáles siempre me han acompañado y acogido muchas personas y distintas instituciones. Por tal motivo, quisiera expresarles mi eterna gratitud y aprecio a todas ellas, sin las cuáles no hubiese podido concluir este proyecto.

Primeramente, agradezco de manera muy especial a mi directora de tesis, Coral Barbas, quién me brindó la oportunidad y, que además, depositó su confianza en mí. Asimismo, durante este doctorado compartió conmigo sus enseñanzas y ejemplos, los cuáles sirvieron como guía a lo largo de estos años. No menos importantes fueron su cariño y apoyo incondicionales, todo un aliciente en los momentos difíciles de este camino.

Agradezco enormemente a la Universidad San Pablo CEU por su hospitalidad, así como a los profesores y personal de esta institución.

Quisiera expresar mi gratitud a la Fundación San Pablo CEU y a EADS-CASA (Airbus Military) por el apoyo económico y por todas las atenciones recibidas, durante mi estancia en España.

Mi más profundo aprecio y gratitud a Elaine Holmes, por permitirme realizar una estancia de investigación en el Imperial College London durante la cual compartió conmigo no sólo sus enseñanzas, sino también su apoyo y cariño. De igual manera, quisiera dar las gracias a los integrantes de su grupo de investigación, que contribuyeron a que la experiencia vivida fuese increíble tanto en el aspecto laboral como extra laboral.

Muchas gracias a Irving Wainer del National Institute of Health (NIH) y a todo su grupo de investigación, por la maravillosa colaboración, entusiasmo y apoyo en este trabajo.

Quiero extender mi aprecio y consideración a Antonia García y Javier Rupérez, quienes siempre me han mostrado su cariño y ayuda; por su amistad, enseñanzas y consejos.

Gracias a Santiago Angulo por todo el conocimiento, apoyo, amistad y afecto brindados a lo largo de estos años.

Mi profundo agradecimiento a Isabel García por todo el gran cariño, apoyo, esfuerzo y participación en este trabajo.

Quisiera externar mi infinita gratitud a James Kinross por el compañerismo, por el esfuerzo y apoyo durante mi estancia en Londres.

Quiero agradecer a todas las compañeras de la sección: Mª Paz Martínez, Mª Ángeles López, Mª Paz Lorenzo y Fernanda Rey-Stolle. Especialmente gracias a nuestra querida Vanesa Alonso por toda su ayuda y afecto.

A todo mis colegas presentes y los que han formado parte del CEMBIO de alguna forma a lo largo de estos cinco años, gracias por todos los momentos tan entrañables que hemos compartido juntos.

Gracias a mis queridos profesores de la Universidad Nacional Autónoma de México la cual ha sido mi primera casa de formación, especialmente a Alma Revilla, Raquel López y Juan José Díaz, por alentarme a buscar esta oportunidad fuera de México.

Finalmente, quisiera expresar mi infinita gratitud a mi familia por todo el tiempo que he dejado de estar con ellos desde que estoy en España, por su entera comprensión, amor y apoyo incondicionales.

INDEX/INDICE

THESIS OUTLINE AND OBJECTIVES

The present doctoral dissertation shows a group of applications in the metabolomics field with a strong emphasis on the utilization of different analytical platforms. Therefore, this thesis has been divided into three sections according to different strategies in metabolomics and analytical methodology: *i*) method development and validation; *ii*) non-targeted metabolomics; and *iii*) targeted analysis. Each one focuses on obtaining different kinds of characteristic information. Taking into account the interest that the analytical techniques have awoken in the scientific field, the general objective of this research work has been to cover most of the analytical techniques in the field of metabolomics that is used in different applications and biological samples.

The main purpose is to show the contribution of analytical techniques, their characteristics and capabilities to achieve the aims of each study. Furthermore, the important aspects of each technique beyond the common aspects mentioned in literature will be discussed in each chapter, including the problems and challenges faced in each specific context. Along with the analytical equipment, the aspects involved with each project – such as sample and data treatment, multivariate and pathway analysis, and biochemical explanation – were integrated as a part of the study.

This thesis is organized as follows: Chapter 1 gives an introduction to all the aspects related to the field of metabolomics, and a description of the workflow stages. The first section (Chapter 2), related to *method development and validation,* describes the characteristics of gas chromatography–mass spectrometry (GC–MS), along with liquid chromatography–mass spectrometry (LC–MS) in the fulfilment of this purpose. This is followed by a section containing the applications of non-*targeted metabolomics,* encompassing three studies focused on the characterization of the metabolic phenotype through the use of different platforms: nuclear magnetic resonance spectroscopy (NMR, Chapter 3); capillary electrophoresis with UV detection (CE–UV) as a complementary technique for NMR (Chapter 4); and liquid chromatography–mass spectrometry (LC–MS, Chapter 5). The third and final section shows the *target analysis* application through the use of capillary electrophoresis with laser induced detector (CE–LIF, Chapter 6).

The particular objectives of each chapter of this dissertation are:

1. To review in an extensive and critical manner the characteristics of capillary electrophoresis (CE) as a tool in metabolomic non-targeted analysis — using either UV detector or coupled to mass spectrometry (MS) — in the analysis of biological samples, and to emphasize the characteristics of CE according to the detection system, and how it affects method development, data treatment and outcomes. This review constitutes part of the introduction in Chapter 1.
2. To develop and compare different extraction protocols for sample extraction of breast milk, in order to achieve by LC–MS a non-target metabolic profile as completely and conveniently as possible in the positive and negative ionization modes (Chapter 2). Then, to optimize and characterize the same extraction solution by using GC–MS, in order to provide wider metabolite coverage. In addition, to validate both methods in LC–MS and GC–MS for selectivity, linearity, accuracy, instrumental precision, method precision (with both standards and samples), limit of quantification (LOQ) and limit of detection (LOD), choosing different metabolites covering the profiles, and with different physicochemical characteristics for each technique. To characterize the breast milk metabolic profiles in LC–QTOF–MS by putative identification through online databases, and in GC–Q–MS by pattern fragmentation and retention index score through the use of Fiehn and NIST libraries. Finally, to apply this methodology to real samples to observe the changes in breast milk composition according to time.
3. To characterize the metabolic phenotype of acute pancreatitis (AP) disease in urine and plasma by NMR. This will then determine the potential of non–target approach by NMR in the diagnosis and prognostic staging of AP to ascertain the clinical utility of this approach in the analysis of acute abdominal pain in the emergency setting (Chapter 3). Furthermore, to compare additional clinical outcomes such as: the aetiology of AP caused by alcohol consumption and cholelithiasis; the AP severity based on the Glasgow severity score; and among the two sub–groups (cholelithiasis and colonic inflammation) inside the heterogeneous control group.
4. To integrate data from two platforms (NMR and CE–UV) to enhance metabolite coverage in the urinary characterization of the slc26a6 knockout mouse model. In addition, to establish the metabolic differences according to genotype and gender between the groups. Also, to apply statistical spectroscopy test to aid peak identification and to establish possible pathway

associations of the significant analytes with other metabolites within the two techniques profile (Chapter 4).

In an annex, to include an additional research work using only CE–UV as a tool for the urinary metabolic characterization of children with diabetes type 1. In this, to observe the nutraceutical effect of an enriched diet at controlled dosage for 1 year in diabetic children, as compared with controls.

5. To employ the general non–target methodology for plasma by LC–QTOF–MS in resistant–treatment patients with bipolar disorder (BD), after ketamine and placebo infusion to resolve the metabolites altered between responders and non–responders to ketamine treatment (Chapter 5). To compare the metabolic differences based on the co–administration of lithium and valproic acid as mood stabilizers in BD patients. To highlight the steps in the data analysis and workflow to get suitable and reliable results for further correct biological interpretation. To innovate, develop and validate the most suitable methodology for target analysis of chiral amino acids in plasma. Afterwards, to validate the method for selectivity, linearity, accuracy, instrumental precision, method precision (both with standards and samples), limit of quantification (LOQ) and limit of detection (LOD), as another objective pursued in this investigation. To characterize the amino acid profile by using standard solutions and test the resolution of L– and D–forms. Finally, to determine in a sensitive and selective manner by CE–LIF these target metabolites in real plasma samples and confirm the quantitation results with values in the literature for BD patients (Chapter 6).
6. To integrate the results and challenges from all these projects into a discussion (Chapter 7), where the strengths, limits and characteristics of these analytical techniques (GC–MS, NMR, CE–UV, LC–MS and CE–LIF) could be confronted.

ESTRUCTURA DE LA TESIS Y OBJETIVOS

La presente tesis doctoral profundiza en los desarrollos metodológicos y aplicaciones en el campo de la metabolómica, con énfasis en el uso de distintos tipos de muestras y diversas plataformas analíticas. Este trabajo se ha dividido en tres secciones de acuerdo a las estrategias que utiliza la metabolómica y a la metodología analítica; *i*) desarrollo de métodos analíticos y validación, *ii*) metabolómica no-dirigida, y *iii*) análisis dirigido, cada uno enfocado en la obtención de distintas características y tipo de información. Teniendo en cuenta las aportaciones que el rápido avance de las técnicas analíticas han realizado en el campo científico, el objetivo general de este trabajo de investigación ha sido el de abarcar la mayoría de las técnicas de análisis dentro del campo de la metabolómica utilizándolas en distintas aplicaciones y muestras biológicas.

El objetivo principal del trabajo es mostrar las contribuciones, características y capacidades de cada una de las técnicas analíticas para alcanzar los propósitos en cada proyecto. De la misma forma los aspectos importantes de cada plataforma, más allá de los aspectos comunes mencionados en la literatura, se han destacado y discutido en cada uno de los capítulos, incluyendo los problemas y desafíos encontrados en cada contexto específico. Además del equipo instrumental utilizado, los aspectos involucrados con cada proyecto como el tratamiento de la muestra y de los datos, el análisis multivariante y de las rutas metabólicas, y la explicación bioquímica se han integrado como parte del estudio.

En una descripción general, esta tesis se encuentra organizada de la siguiente forma; el Capítulo 1 muestra una introducción de todos los aspectos relacionados con la metabolómica y una descripción completa de cada una de las etapas del proceso de trabajo en este campo. El primer bloque temático de este trabajo (Capítulo 2) está relacionado con el *desarrollo y validación de métodos analíticos,* y exhibe las características de la cromatografía de gases acoplada a la espectrometría de masas (GC-MS) que, en conjunto con la cromatografía de líquidos acoplada a espectrometría de masas (LC-MS), para obtener métodos analíticos robustos y de gran valor para la caracterización del perfil de fluidos biológicos. La segunda sección del trabajo contiene las aplicaciones de la metabolómica en el *análisis no-dirigido* que incluye tres estudios que se enfocaron en la caracterización metabólica del fenotipo a través de la aplicación de distintas plataformas analíticas; la espectroscopía de resonancia magnética nuclear (RMN,

Capítulo 3), la electroforesis capilar con detección UV (CE-UV) como una técnica complementaria para RMN (Capítulo 4) y la cromatografía de líquidos acoplada a espectrometría de masas (LC-MS, Capítulo 5). Finalmente la tercera sección expone una aplicación del *análisis dirigido* a través del uso de la electroforesis capilar con detector de fluorescencia inducida por láser (CE-LIF, Capítulo 6).

Los objetivos particulares de cada capítulo para esta tesis son:

1. Revisar de manera extensa y crítica las características de CE como una herramienta en el análisis no dirigido– utilizando como detector el UV o acoplada a espectrometría de masas (MS) – en el análisis de muestras biológicas. Enfatizar las características de la CE de acuerdo con el sistema de detección y cómo este afecta el desarrollo del método analítico, el tratamiento de datos y los resultados obtenidos. Este artículo formó parte de la introducción, en el Capítulo 1.
2. Desarrollar y comparar diferentes protocolos de extracción para la preparación de muestra de leche humana con el fin de lograr un único extracto, compatible con LC-MS y GC-MS y que proporcione el perfil más amplio posible de compuestos. Obtener un perfil metabólico no-dirigido en LC-MS lo más completo y conveniente como sea posible en los modos de ionización positivo y negativo (Capítulo 2). Realizar la optimización del método para la caracterización del mismo extracto mediante el uso de GC-MS. Además, validar ambos métodos en LC-MS y GC-MS en cuanto a selectividad, linealidad, exactitud, precisión precisión, límite de cuantificación y detección, escogiendo distintos metabolitos cubriendo los perfiles y con diferentes características fisicoquímicas de acuerdo a cada técnica. Caracterizar los perfiles metabólicos de la leche materna en LC-QTOF-MS a través de la identificación tentativa en las bases de datos y en GC-MS-Q a través del patrón de fragmentación y la puntuación del índice de retención (RI) de acuerdo con las librerías de Fiehn y NIST. Finalmente aplicar esta metodología a muestras reales para observar los cambios en la composición de la leche materna con el tiempo.
3. Caracterizar el fenotipo metabólico de la enfermedad de pancreatitis aguda (AP) en orina y en plasma por espectroscopía de resonancia magnética nuclear de protón (^{1}H RMN). Para ello, determinar el potencial del enfoque del análisis no-dirigido por RMN en el diagnóstico y la etapa de pronóstico de la AP para determinar la utilidad clínica de este enfoque en el análisis del dolor abdominal agudo en la unidad de emergencias (Capítulo 3). Además, comparar otros estados clínicos adicionales tales como: la etiología de la AP causada por el consumo de alcohol y por

colelitiasis, la severidad de la AP en base a la gravedad del parámetro de Glasgow, y entre los dos subgrupos (colelitiasis e inflamación del colon) dentro del grupo heterogéneo de controles.

4. Integrar los datos a partir de dos plataformas analíticas (RMN y CE-UV) para aumentar la cantidad de metabolitos en la caracterización de orina del modelo de ratón knockout slc26a6. Además, establecer las diferencias metabólicas de acuerdo al genotipo y al sexo entre los grupos. También, aplicar la correlación estadística espectroscópica para ayudar a la identificación de compuestos en CE y establecer posibles asociaciones de los metabolitos más significativos con otros en el perfil combinando las dos técnicas (Capítulo 4).

   En un anexo, incluir un trabajo de investigación adicional utilizando únicamente CE-UV como herramienta para la caracterización metabólica en orina de niños con diabetes tipo 1. En este, observar el efecto de una dieta nutracéutica enriquecida en una dosis controlada durante 1 año en niños diabéticos comparados con controles.

5. Emplear la metodología general del análisis no-dirigido para la obtención del perfil global de plasma por LC-QTOF-MS en pacientes resistentes al tratamiento con desorden bipolar (BD) después de haber sido infundidos con ketamina y placebo para encontrar los metabolitos alterados entre pacientes que responden y los que no responden al tratamiento con ketamina (Capítulo 5). Comparar las diferencias metabólicas basadas en la co-administración de litio y ácido valproico como estabilizadores del estado de ánimo en pacientes con BD. Resaltar los pasos del análisis de datos y del flujo de trabajo para obtener resultados fiables para su correcta interpretación biológica.

6. Innovar, desarrollar y validar la metodología más adecuada para el análisis dirigido de aminoácidos quirales en plasma. Después, validar el método analítico para selectividad, linealidad, exactitud, precisión instrumental y del método (tanto con patrones y muestras), límite de cuantificación y detección, como parte de otro objetivo perseguido en esta investigación. Caracterizar el perfil de aminoácidos mediante el uso de soluciones estándar y probar si la metodología es capaz de la resolución de las formas L- y D- de los aminoácidos. Por último, determinar de una manera sensible y selectiva por CE-LIF estos metabolitos diana en muestras de plasma reales y comparar los resultados de la cuantificación con valores publicados en la literatura para pacientes con BD (Capítulo 6).

7. Integrar los resultados y desafíos de todos estos proyectos en una discusión (Capítulo 7), donde las fortalezas, los límites y las características de cada una de las técnicas analíticas (GC-MS, RMN, CE-UV, LC-MS y CE-LIF) puedan ser confrontadas.

Resumen

Años atrás el pensamiento de tener un conocimiento profundo de los procesos biomoleculares y del desarrollo de nuevas pruebas de diagnóstico que pudieran conllevar a un tratamiento médico personalizado era inimaginable. La metabolómica fue la última de las ciencias "*-ómicas*" que surgió como disciplina después de la realización del proyecto genoma humano. Cuando el proyecto del genoma humano había concluido se pensaba que la expresión génica explicaría el fenotipo de los seres vivos, por lo que billones de dólares y un intenso trabajo fueron puestos para construir lo que sería la gran industria de las ciencias "*-ómicas*". Estas ciencias "*-ómicas*" empezaron con la medición de los genes (genómica), la expresión de los mismos (transcriptómica) y con la medición también de los cambios en la traducción proteica (proteómica). Sin embargo, a pesar del avance en estas disciplinas no se encontró suficiente información para entender el fenotipo de un sistema vivo. En aquel tiempo el descubrimiento genético había excedido la habilidad de entender completamente las funciones génicas y su relación con el fenotipo observado.

La metabolómica entonces apareció como respuesta a la interrogante de como evaluar el fenotipo. Está ciencia fue la última en desarrollarse debido a dos factores limitantes; el primero basado en el cambio de pensamiento sobre el dogma central de la biología molecular y el segundo enfocado a la necesidad de nuevas herramientas analíticas de alto rendimiento y de sistemas informáticos y estadísticos apropiados. El nuevo dogma central de la biología molecular reconocería pues al metabolismo como un sistema de rutas metabólicas que se encuentran íntimamente interconectadas en redes dinámicas y complejas, donde el flujo de información de [gen – transcripción – proteína – metabolito] estaría en constante retroalimentación. En este nuevo concepto sin embargo, los metabolitos continuarían siendo los productos finales del metabolismo y por lo tanto serían capaces de caracterizar el fenotipo en un organismo. Este interés entonces de medir los metabolitos para caracterizar el fenotipo llevó a Oliver S. G. en 1998 a introducir el término "*metaboloma*" y lo uso para describir la cantidad total de metabolitos de bajo peso molecular en un organismo. Luego el concepto de "*metabonómica*" sería acuñado por Nicholson, J.K. en 1999 para definirlo como "*la medición cuantitativa múlti-paramétrica de la respuesta metabólica como respuesta de los sistemas vivos a estímulos patofisiológicos o modificaciones genéticas*". Algún tiempo después Fiehn, O., en 2000 introduciría la palabra "*metabolómica*" para citar el

"*estudio imparcial de todas las moléculas de bajo peso molecular o metabolitos en un fluido biológico, célula, tejido, órgano u organismo*". Sin embargo en la actualidad ambas palabras metabolómica y metabonómica son usadas indistintamente.

Los metabolitos son pues el resultado de la combinación de factores genéticos y ambientales tales como la dieta, la edad, la etnicidad, el estilo de vida y la actividad de la flora intestinal. Por ello cuando cualquier nivel fisiológico en un organismo es alterado, e.g. los niveles de alguna proteína en respuesta a alguna señal ambiental o patológica el organismo va a tratar de mantener el estado de equilibrio u homeostasis frente al cambio. En el caso de organismos complejos como los seres humanos, existen múltiples mecanismos que permiten tener flexibilidad para manejar las perturbaciones fisiológicas y del metabolismo. Esta flexibilidad caracteriza que tan propenso es un organismo vivo de mantenerse funcionando efectivamente frente a factores de estrés internos y externos. Sin embargo cuando un sistema ya no es capaz de responder efectivamente, esté desarrolla un estado de enfermedad y como consecuencia las rutas metabólicas son perturbadas generando un aumento o disminución de un grupo de metabolitos. Este proceso tiene como resultado que los cambios en los metabolitos sean amplificados dentro del metaboloma dando así una descripción altamente sensible del estado fisiológico de un ser vivo.

Sin embargo, la medición de todos los metabolitos en un sistema e inclusive en una muestra es un objetivo que está lejos de ser posible. Esto debido principalmente a dos razones; la primera es que la cantidad de metabolitos es muy grande, y la segunda es que poseen propiedades físicas y químicas muy diversas, y que se encuentran en distinta magnitud de concentración en el organismo. Debido a esto, la metabolómica ha abordado ciertas estrategias para medir estos metabolitos, las cuales siguen aún evolucionando:

1. Análisis dirigido (*target analysis*). Esta estrategia precisa y cuantitativa pretende el análisis de uno o varios metabolitos los cuales generalmente están relacionados en alguna ruta metabólica. Esta estrategia puede estar derivada del conocimiento biológico previo de algún estudio metabólomico (e.g. la validación de potenciales biomarcadores) o cuando los límites de detección son muy bajos. Este tipo de análisis también es usado como parte de otros análisis dentro del flujo de trabajo (e.g. en la identificación de metabolitos por masas en tándem). Finalmente esta

estrategia ha sido aplicada tradicionalmente en la química analítica tradicional y en ensayos bioquímicos en laboratorios clínicos.

2. Perfiles metabólicos. Inicialmente esta estrategia estuvo enfocada en medir grupos de metabolitos específicos, e.g. aminoácidos, ácidos grasos libres, etc. Ahora este término no es más que un sinónimo del análisis no-dirigido.
3. Análisis no-dirigido (*Non-target analysis*). Es la determinación simultánea y no sesgada de múltiples metabolitos suficientes para representar el estado metabólico de un organismo mediante una única medición.
4. Pisada metabólica (*footprinting*). También llamada exo-metaboloma o secretoma, su propósito es el de medir los metabolitos extra celulares, es decir los componentes del medio en un cultivo celular.

De estas cuatro estrategias el análisis no-dirigido es la estrategia más ampliamente utilizada en estudios metabolómicos. Inicialmente en esta estrategia no hay interés en ningún compuesto en particular sino en comparar los patrones o perfiles metabólicos y detectar las señales que cambian estadísticamente entre los grupos de estudio. Aquí sin importar la variabilidad singular de cada análisis no-dirigido algunos metabolitos característicos van a encontrarse alterados como respuesta a un estado fisiológico en particular.

Aunque hay una amplia cantidad de ejemplos que siguen aumentando y que muestran el desarrollo y la aplicación del análisis no-dirigido dentro de la biología de sistemas; e.g. en pruebas de toxicología, en el diagnóstico y/o pronóstico de enfermedad, y en la eficacia de fármacos, esta estrategia sigue normalmente un flujo de trabajo común dentro del área de metabolómica. Este empieza estableciendo el modelo biológico de acuerdo a los objetivos perseguidos, el presupuesto económico y los recursos tecnológicos con los que se cuentan. Lo siguiente llevar a cabo el estudio biológico para recolectar las muestras biológicas siguiendo un método estándar y manteniéndolas en las condiciones apropiadas hasta el momento de su análisis. El protocolo de preparación de muestras va a depender de la técnica analítica a utilizar y del enfoque metabolómico que se persiga (análisis dirigido o no-dirigido).

Después de su medición en una plataforma analítica, los datos adquiridos son tratados de acuerdo al tipo de datos obtenidos por la técnica analítica, e.g. para RMN y CE-UV los datos obtenidos son de

tipo 2D (señal analítica, tiempo o ppm) mientras que para cualquier técnica analítica acoplada a MS los datos son tipo 3D (masa, tiempo y abundancia). Para el tratamiento de datos y el estudio de las diferencias estadísticas entre dos grupos se utilizan herramientas multivariantes y quimiometría. Para datos tipo 2D el tratamiento de datos consta en general de la reducción del ruido, eliminación de la línea base, alineamiento de los perfiles y normalización cuando sea el caso. En caso de los datos tipo 3D y que se refieren a los obtenidos por MS, el tratamiento de los datos básicamente consiste en la detección de entidades químicas (*features*), alineamiento, normalización y filtración de los datos. Sin embargo el tratamiento de datos detallado va a depender del tipo de ionización y del analizador de masas.

Para el análisis estadístico de los datos existen dos estrategias para abordar este propósito: a través del análisis univariante y multivariante. El análisis univariante es una asociación simple de cada variable con la clase del caso de estudio, esta estrategia se basa en modelos de regresión lineal que pueden ser paramétricos (e.g. prueba *t*- de Student para dos grupos o ANOVA para tres o más grupos), como no-paramétricos (e.g. prueba de la U de Mann-Withney para dos grupos o Kruskal-Wallis para 3 grupos o más). Esta estrategia sin embargo, considera cada variable como independiente e ignora la correlación o efectos entre las variables por lo que aunque contribuye al proceso estadístico es en realidad una herramienta complementaria del análisis multivariante.

Por otra parte, el análisis multivariante es usado para analizar múltiples variables al mismo tiempo a través de dos tipos de métodos: el no-supervisado y el supervisado. En el análisis no supervisado, el software no utiliza ninguna información *a priori* de los datos y busca reducir la variabilidad de cada una de las muestras a través de las variables que puedan explicar alguna agrupación de las muestras. Esto permite al investigador observar tendencias o agrupaciones de los datos, permitiendo así identificar muestras que estén fuera de tendencia (*outliers*), y mostrando una visión general de todas las muestras en el estudio. Entre los principales métodos el análisis de componentes principales (PCA) es el principal. En los métodos supervisados la información del grupo al que pertenece cada muestra es utilizada. Ente los métodos más utilizados el análisis discriminante de la regresión de mínimos cuadrados parciales (PLS-DA) y de regresión ortogonal (OPLS-DA) son los más comunes. Los métodos de PLS-DA pueden extraer las variables de la matriz de datos (*X*) que estén altamente correlacionadas y que cavarían con la clase (*Y*). El resultado de este método genera un modelo que basa la separación de los grupos de estudio en más de un vector de variables llamados componentes (*t*) donde el

primero corresponde a la clase a la que pertenecen cada muestra. En este modelo sin embargo el proceso para interpretar cuales son los metabolitos que producen la separación debida únicamente a la clase no es del todo directo y/o especifico. Por lo tanto los modelos de OPLS-DA, incluyen toda la información relevante que explique la separación de los grupos en una sola componente predictiva (que corresponde a la clase) de la variación no predictiva también llamada ortogonal. Por tanto en general, los modelos de PLS-DA son usados para clasificar las muestras de acuerdo con la clase, mientras que los modelos de OPLS-DA son usados para la búsqueda de biomarcadores.

Una vez obtenidos los compuestos significativos el objetivo es poder identificarlos y con ellos finalmente poder obtener de la mejor manera una explicación de la desregulación de estos, a través de alguna de las rutas metabólicas y si es posible por bajo una hipótesis biológica.

Dentro del flujo de trabajo todos los pasos son importantes y el análisis químico es precisamente uno de los pasos críticos. Fue de hecho gracias al avance tecnológico de los últimos años concretamente en el desarrollo de técnicas analíticas robustas y de alto rendimiento, que junto con las herramientas de quimiometría y el análisis de reconocimiento de patrones que la investigación en metabolómica despego y creció en el campo de la investigación. Por lo que las plataformas analíticas tienen un rol primordial en los estudios metabolómicos, sus características y aplicaciones fueron el principal objetivo de este trabajo doctoral y por lo tanto la descripción completa de cada una ha ocupado el curso de los siguientes capítulos de este trabajo.

El presente trabajo ha abarcado y mostrado la mayoría de las plataformas analíticas aplicadas en el campo de la metabolómica. El objetivo de esta tesis ha sido dar una visión global de cada una de las técnicas analíticas a través de una aplicación biológica que ha concluido con una publicación. Cada uno de los capítulos está dedicado a una plataforma analítica distinta. Al inicio de cada uno de ellos se encuentra citado en un resumen los fundamentos de cada técnica analítica, las ventajas, los puntos débiles, los desafíos y problemas que actualmente enfrenta cada una. Junto a estos parámetros, aspectos técnicos de los equipos, capacidades y la versatilidad de cada instrumento es señalado. En un sentido más amplio y extenso de cada capítulo y plataforma analítica:

El Capítulo 2 describe el desarrollo y validación del análisis no-dirigido para la leche materna. El objetivo principal de este proyecto fue el desarrollar un procedimiento de extracción común para obtener la mayor cantidad de metabolitos en la leche materna (BM) la cual permitiera su caracterización en dos técnicas analíticas distintas (LC–MS y GC–MS) aumentando así la cobertura del metaboloma. La BM representa en teoría el único alimento para un recién nacido y los cambios en su composición se han asociado a producir efectos secundarios en el desarrollo del recién nacido a largo plazo. Este fluido biológico se encuentra constituido en su mayoría por lípidos intrínsecamente envasados en forma de glóbulos, lo que conduce normalmente a la formación de dos fases y consecuentemente que alguna manipulación no se puede evitar. El método de extracción optimizado y seleccionado en este capítulo utiliza metil *ter*-butil éter (MTBE) y metanol (1:1) con lo que se obtiene un sistema estable de una sola fase de la muestra de leche. Este importante aspecto no solo simplifica el tratamiento de la muestra, reduce el tiempo y la manipulación de la muestra, evitando el mecanismo de partición de los metabolitos cuando se forman de dos fases en el sistema.

GC-Q-MS fue la primera técnica analítica utilizada en acoplamiento y fue descrita como la plataforma estándar para la obtención del perfil metabólomico no-dirgido en distintos tipos de muestras. Esta técnica puede considerarse como una técnica altamente selectiva pues solo aquellos metabolitos que sean volátiles y térmicamente estables (por naturaleza o después de un proceso de derivatización) pueden ser analizados. La GC es una técnica de alta resolución y reproducibilidad, que acoplada al equipo de MS permite una alta sensibilidad de detección. Además debido a que generalmente el equipo de MS acoplado a GC utiliza como fuente de ionización el del tipo de impacto electrónico (EI), esto permite la creación de librerías espectrales que después son usadas en el proceso de identificación haciendo que este sea más rápido y confiable.

Para esta técnica su mayor inconveniente es precisamente el proceso de derivatización, al cual la mayoría de las muestras biológicas tienen que ser sometidas. Este proceso tiene un alto riesgo de sesgo o de introducir variabilidad instrumental al análisis, además generalmente produce múltiples productos de la derivatización y más de un pico para varios metabolitos. Esto ha hecho que varios autores consideren a la GC-MS como una técnica de bajo rendimiento.

Debido a todos los aspectos de la GC-Q-MS, esta fue una de las técnicas utilizadas para la caracterización metabólica de la BM, utilizando el método de bloqueo del tiempo de retención (RTL) llamado "*Fiehn GC-MS metabolomics*", el cual ajusta la presión del gas portador de la fase móvil permitiendo obtener tiempos de retención (RT) muy reproducibles. Esto permite como consecuencia también, que la identificación de los compuestos en el perfil sea hecha no solo por el patrón de fragmentación, sino por el tiempo de retención a través del índice de retención, RI. Como resultado de la aplicación de este método 60 metabolitos fueron identificados, entre ellos aminoácidos, ácidos orgánicos, ácidos grasos libres (FFA), azucares (hexosas y pentosas), intermediarios del ciclo de Krebs (TCA), colesterol y disacáridos. Sin embargo debido a la cantidad tan alta de disacáridos en el extracto, el rango de tiempo de retención correspondiente a esta zona (minuto 23 al 26) de azucares el filamento de la fuente de ionización fue apagado para prevenir que el filamento se ensuciara y con ello evitar la disminución de la señal analítica.

Hasta la fecha no existe ninguna guía establecida de como validar los métodos analíticos no-dirigidos en metabolómica, por lo que se optó por utilizar la guía habitual para métodos bioanalíticos, para esto se seleccionaron 17 metabolitos de diferente clase bioquímica y a ser posible cubriendo todos los RTs de los cromatogramas en GC-MS y LC-MS. La validación demostró ser válida para los siguientes parámetros: selectividad, linealidad, exactitud, precisión (reproducibilidad y precisión intermedia), límite de cuantificación y detección. El tratamiento de datos cuando se combinan ambas plataformas analíticas tiene que ser independiente debido a que los quipos de MS son totalmente distintos, tienen una fuente de ionización y analizador de masas son diferentes por lo que el tratamiento de datos es también distinto. Con esto ambas técnicas dan una información distinta y complementaria, sin dejar de tener algunos compuestos comunes entre ambas.

El Capítulo 3 es un estudio que muestra la caracterización metabolómica de plasma y orina de pacientes con pancreatitis aguda (AP) y pacientes con dolor abdominal agudo (controles) en la unidad de emergencias del hospital. Este estudio cubrió un objetivo muy ambicioso como es la aplicación de la espectroscopia RMN en el ámbito clínico. En él, tres escenarios clínicos fueron explorados; 1) la comparación entre AP y controles, para la búsqueda de biomarcadores específicos para el diagnóstico de AP, 2) la diferencia de acuerdo a la etiología de la enfermedad; alcoholismo o cálculos biliares, y 3) de acuerdo a la severidad de AP, valorada mediante el test de Glasgow comparando así los pacientes con AP

leve y moderada. Adicionalmente se logró comparar dentro del grupo de los controles (caracterizado por la presencia de dolor abdominal agudo) la comparación de los grupos de pacientes con inflamación de colon y pacientes con cálculos biliares.

Como resumen de la técnica empleada: la espectroscopía de resonancia magnética nuclear (RMN) es una técnica no destructiva, sin sesgo, robusta y altamente reproducible, que provee un perfil metabólico rápido, con buena resolución y que requiere una mínima preparación de muestra adecuada para muestras líquidas y sólidas. Esta tecnología considerada como de alto rendimiento permite medir hasta 400 muestras en 24 h. El espectro obtenido es representativo del estado metabólico del paciente adquirido en una sola lectura. Esta técnica que depende del núcleo seleccionado (típicamente ^{1}H) muestra todas las moléculas suficientemente abundantes que contengan este núcleo (desde μM). Una de las principales ventajas de esta técnica es que permite obtener información estructural detallada de los metabolitos en las muestras biológicas, que para su identificación, la cual se puede ayudar de bases de datos que contengan los espectros de RMN de los estándares e.g., HMDB.

Dentro de las desventajas que tiene esta técnica la mayor es su baja sensibilidad y límite de detección. Adicionalmente, los espectros de RMN presentan generalmente cientos de compuestos que presentan distintos grados de solapamiento. Esto hace que el tratamiento de los datos sea altamente complejo e.g., el solapamiento obstaculiza la asignación completa de los picos en los espectros de ^{1}H RMN en la mayoría de compuestos. El problema de asignación de las señales en RMN aumenta conforme los metabolitos se acercan al límite de detección. Más aún en muestras biológicas como por ejemplo plasma, donde hay la presencia de macromoléculas como proteínas y lípidos, el espectro de RMN se caracteriza por señales gruesas y anchas que enmascaran la presencia de metabolitos pequeños conduciendo a que, su detección y cuantificación sean difíciles. Para solucionar este problema se creó una secuencia en RMN llamada "*spin-eco sequence Carr-Purcell-Meiboom-Gill (CPMG)*" la cual selecciona únicamente las señales provenientes de los metabolitos.

Dentro de los problemas específicos encarados durante la realización del proyecto por RMN se observó el alto impacto en el perfil de RMN debido a la presencia de xenobióticos (e.g. fármacos o excipientes), o debido al alto contenido de proteínas en la muestra debidas a la patología y que no son usuales de encontrar en orina. Ambas cosas pueden causar sesgo o malinterpretación de los resultados. En

el caso de las muestras de orina que presentaron una gran concentración de proteína, estas muestras fueron descartadas del estudio, disminuyendo así el número de pacientes en los grupos de comparación. En el caso de los xenobióticos su presencia muchas veces en el perfil puede no impactar el análisis de datos sin embargo cuando lo hace, estos deben de ser quitados de los datos originales para permitir obtener resultados reales. Como consecuencia de cortar las señales de los xenobióticos en los datos originales, todos los compuestos que se encuentren muy cercanos o sobrepuestos en estas regiones van a quedar fuera del estudio.

El Capítulo 4, es un estudio basado en la caracterización del perfil metabólico en orina de ratones knockout slc26a6 en comparación con los controles. Este estudio se llevó a cabo a través de la combinación de dos plataformas analíticas complementarias (RMN + CE-UV). En este estudio, claras diferencias en los perfiles de orina fueros observados de acuerdo al genotipo y al sexo. Sin embargo, la combinación de los perfiles de ambas técnicas permitió el estudio de correlaciones de los dos picos más discriminantes del estudio con todos los compuestos provenientes de ambas técnicas. Más aún a través de este estudio de correlaciones se logró identificar tentativamente uno de los compuestos más significativos en CE-UV, a través del espectro de RMN. La identificación sin embargo no se logró corroborar porque no existe el reactivo comercial.

La CE es una técnica de separación que se caracteriza por su alta eficiencia y resolución de un rango diverso de compuestos químicos, particularmente útil para separar los compuestos polares y con carga. La CE tiene distintas características entre ellas esta su gran versatilidad y selectividad, ya que es capaz de trabajar en varios modos de separación en un tiempo corto de análisis, de ellos la electroforesis capilar zonal (CZE) y la cromatografía micelar electroquinética (MEKC), son las más comunes incluso en metabolómica. Otras ventajas de la CE es que requiere poco volumen de muestra, cantidades bajas de soluciones tampón y puede utilizar distintos sistemas de detección (absorción UV-vis, LIF, MS). Adicionalmente después de cada análisis el capilar puede ser fácilmente lavado reduciendo así la probabilidad de obstrucción. Uno de los aspectos importantes de la CE es el bajo consumo de solventes orgánicos (inclusive ninguno), que la hace ser considerada como una "*tecnología verde*", además de su bajo costo comparado con todas las técnicas actuales usadas en el área. Hay que resaltar que la CE tiene la habilidad de separar matrices complejas con apenas una simple preparación de muestra, e.g. la orina puede ser inyectada directamente con una simple dilución con agua.

Sin embargo el mayor inconveniente es su baja sensibilidad y poder de identificación cuando el detector empleado es el de absorción UV-vis. Este sistema de detección es poco sensible debido a la baja cantidad de volumen de muestra que es introducido en el capilar (nanolitros) y la corta longitud de paso óptico. Además, el sistema de detección no es universal así que los compuestos sin cromóforos (e.g., carbohidratos) no son detectados. El problema de la identificación en CE-UV se encuentra debida a que el detector no aporta información estructural, por lo que la identificación de los compuestos en el electroferograma se lleva a cabo añadiendo el estándar puro a la muestra, este proceso se realiza en forma de ensayo y error con estándares puros de los compuestos esperados en el perfil. La asignación completa se realiza por comparación de los tiempos de migración, espectro de absorción y por el crecimiento en el pico cuando el estándar es añadido a la muestra. Como punto débil adicional de la CE se encuentra la robustez de los tiempos de migración.

La caracterización metabolómica de los biofluidos fue llevada a cabo a través de la aplicación de dos metodologías previamente establecidas en CE-UV. Estas metodologías trabajan en polaridades opuestas y explotan de esta manera dos distintos modos de separación distintos en CE, uno por MEKC (en polaridad normal) y otro en CZE (en polaridad invertida). De aquí la metodología en MEKC permite el análisis de cationes y moléculas neutras, seguidas de aniones grandes. Por el contrario, el método por CZE permite la detección básicamente de iones pequeños, como ácidos orgánicos de cadena corta. Además del rango más amplio de metabolitos detectados uniendo ambas metodologías, estos métodos muestran varias ventajas; los dos son fáciles de aplicar, la preparación de las soluciones tampones son fáciles y baratas, el tiempo de análisis es corto y hay un desplazamiento de los picos relativamente bajo (debido a que la corriente generada por el capilar es relativamente baja de -70μA para CZE y 150μA para CD-MECK) y que puede ser corregido mediante un algoritmo matemático en Matlab (e.g. el algoritmo Correlation Optimized Warping (COW)).

Para llevar a cabo la caracterización metabólica de orina se adoptó una estrategia dual combinando los espectros de RMN y de CE-UV debido a que ambas técnicas producen matrices de datos tipo 2D, por lo que el tratamiento de datos es muy parecido y permite su integración en un solo perfil. Además mientras que la RMN genera un perfil robusto y amplio de un gran rango de compuestos de bajo peso molecular, CE-UV proporciona datos de especies polares y iónicas incluyendo oxalato (el cual no es detectado por RMN), cloruros y nitratos. Es importante resaltar el hecho del volumen tan pequeño que

requiere la CE-UV para trabajar adecuadamente, el cual para esta aplicación en orina de ratón fueron únicamente 15µL en total.

La combinación y el tratamiento de datos de RMN y CE-UV son mucho más fáciles que la combinación con cualquier otra técnica acoplada a la MS. Las diferencias entre ambas técnicas usadas, en términos de contribución a todo el estudio pueden ser explicadas por la distinta capacidad de cada técnica para detectar distintos tipos de compuestos dentro de una muestra biológica. Para este estudio e.g. el oxalato y los nitratos son detectables en la CE-UV pero no en el espectro de RMN, mientras que la taurina y las metilaminas son más fácilmente observables en los perfiles de RMN.

El Capítulo 5, muestra la caracterización metabolómica de plasma de pacientes resistentes a tratamiento que presentan trastorno bipolar (BD) los cuales fueron sometidos a un nuevo fármaco, en un estudio cruzado con una infusión de ketamina y placebo. Estos pacientes contaban además con un medicamento alterno como estabilizador de ánimo; litio o ácido valproico. Las muestras fueron analizadas por LC-QTOF-MS y el objetivo de este estudio fue el emplear una estrategia de análisis no-dirigido para identificar a los potenciales marcadores de la respuesta a ketamina de los pacientes que no respondieron al nuevo procedimiento médico. Los resultados demostraron que los perfiles metabolómicos fueron significativamente distintos entre los pacientes que tomaban litio del grupo mantenido con ácido valproico independiente de la respuesta al tratamiento. Por lo que los grupos fueron estudiados por separado. En los pacientes bajo tratamiento con litio, 18 marcadores fueron identificados correlacionados con la respuesta a ketamina. Estos compuestos indicaron que posiblemente las diferencias entre los pacientes que responden al tratamiento con ketamina de los que no lo hacen se deben a alteraciones en la β-oxidación de los ácidos grasos en la mitocondria. Con esto se demostró el alto potencial de LC-QTOF-MS como una herramienta en la predicción de la respuesta y en la individualización de la terapia con ketamina.

En general, la LC-MS es una de las técnicas analíticas que más se emplean en la actualidad en el área de metabolómica tanto para el análisis cuantitativo dirigido como para la obtención de los perfiles del análisis no-dirigido. Esta es una técnica de alta sensibilidad, exacta y reproducible, capaz de detectar una gran cantidad de clases de metabolitos dentro de un amplio rango dinámico de masa (generalmente de 100 a 1000 Da). Si se tiene cuidado está técnica analítica no solo provee un fenotipo metabólico detallado sino

que permite obtener información estructural de los compuestos a través de la distribución isotópica de su espectro de masas (MS) o a través de su perfil de fragmentación de MS/MS. Esta última característica permite que esta técnica sea muy selectiva permitiendo aislar los compuestos de interés. Adicionalmente está técnica si se trabaja con cuidado es muy robusta y reproducible. La parte de MS concede a esta técnica la posibilidad de analizar en modo positivo y negativo, y el acoplamiento con LC permite una gran resolución de los compuestos.

El trabajo con esta técnica no es fácil, está requiere de cierta experiencia para poder llevar a cabo correctamente los experimentos. En este equipo es importante estar checando continuamente los parámetros del sistema analítico y del desempeño (e.g. curva de presión, intensidad de señal, forma del perfil, etc.). Los datos generados por el equipo de LC-MS son en general complejos, además de ser datos tipo 3D (masa, tiempo y abundancia), son datos caracterizados por presentar ruido y señales aleatorias por lo que uno de sus mayores inconvenientes es el tratamiento de los datos, ya que es importante trabajar con información fiable que luego pueda ser comparada estadísticamente. En este sentido una manera de obtener los compuestos fiables en un experimento de LC-MS (y en general en cualquier técnica que utilice MS) es mediante el uso de muestras llamadas controles de calidad ("*quality control*"). Estos controles de calidad son una mezcla de todas las muestras en el análisis y son medidas varias veces durante el experimento para poder monitorizar la estabilidad y desempeño instrumental.

Entre los inconvenientes de la LC-MS para un conjunto de muestras grande, están su bajo desempeño y estabilidad analítica aunados a los problemas en la eficiencia del proceso de ionización (e.g. ESI tiene un gran potencial de presentar supresión iónica). La LC-MS es además una técnica destructiva por lo cual se vuelve complicada para el análisis de muestras valiosas, además de que los protocolos de preparación de muestra son laboriosos, requiere además para la cuantificación de metabolitos de un estándar interno, y finalmente esta técnica tiene un alto costo.

Adicionalmente, dentro de los problemas enfrentados al trabajar con esta técnica están el tratamiento de datos (previamente mencionado) y la identificación de los compuestos significativos. Para el tratamiento de los datos de MS no hay un procedimiento estándar y aspectos como: el manejo de los datos faltantes, un método de normalización para datos de MS y la transformación logarítmica para el cálculo estadístico univariante, son aspectos que no están del todo claros.

El proceso de identificación por LC-MS es complejo y por lo tanto lleva tiempo, además este enfrenta diversos desafíos, lo que conlleva a que la identificación de un compuesto tentativo no pueda llevarse a cabo o inclusive, que un compuesto desconocido ("*unknown*") no pueda llegar a ser identificado. Como ejemplo el artículo del Capítulo 5, muestra que 165 metabolitos fueron significativos entre el grupo de pacientes que respondieron al tratamiento del grupo que no lo hizo (todos bajo la co-medicación con litio), sin embargo solo 18 compuestos pudieron ser identificados después de los experimentos de MS/MS, y de estos solo un reactivo estándar estaba disponible.

Para este estudio el total de comparaciones entre los distintos grupos fueron 6, este hecho dificultó que se pudiera llevar a cabo el proceso de identificación por experimentos de MS/MS de todos los metabolitos significativos, por lo que solo se enfocó en identificar los metabolitos de la comparación de mayor interés científico.

Cada una de estas plataformas analíticas (RMN, CE-UV y LC-QTOF-MS) tiene un gran valor según las características y capacidades, todas ellas han demostrado su contribución en el área de la metabolómica, siguiendo la estrategia del análisis no-dirigido. Es importante resaltar que dentro de las técnicas analíticas actuales que también se están empleando para obtener el perfil metabólico no-dirigido, la CE-MS y la GC-MS son herramientas ampliamente utilizadas (la GC-MS fue de hecho utilizada en el Capítulo 2 de este trabajo para la caracterización de leche materna).

Finalmente el Capítulo 6 muestra una aplicación de análisis dirigido. La cuantificación del perfil de aminoácidos en plasma en pacientes con alguna condición neuropsiquiátrica es de suma importancia. Durante el paso del tiempo diversas enfermedades como: la manía, esquizofrenia, paranoia, y por supuesto el desorden bipolar (BD), se han asociado a cambios en las concentraciones de aminoácidos en plasma. Sin embargo hasta ahora los aminoácidos no han sido estudiados en pacientes con BD tratados con una infusión de ketamina. Más aún no se ha comparado si hay algún aminoácido que pueda estar relacionado con la respuesta o la falta de ella al tratamiento con ketamina de estos pacientes. Este estudio tuvo como objetivo optimizar y validar un método analítico lo suficientemente sensible y selectivo para determinar los aminoácidos en plasma y con ello aplicarlo para cuantificar estos en pacientes con BD.

La metodología en un análisis dirigido va a depender siempre de las características físico-químicas de los compuestos de interés. En este caso los aminoácidos son compuestos muy polares por lo que es difícil extraerlos usando solventes polares para su posterior análisis en GC, no son retenidos cuando se intenta medirlos por RP-LC y no presentan cromóforo por lo que su detección también es difícil, todo esto hace que su determinación represente un reto analítico. De entre las técnicas que se han empleado para determinar estos compuestos, la CE-LIF representa una buena opción. Esta técnica está caracterizada por la combinación de la CE que permite una separación de alta resolución y el LIF como uno de los detectores más sensibles y altamente selectivos pues solo es capaz de detectar fluorescencia emitida por las moléculas ya sea natural o mediante la unión de los compuestos de interés con un fluoróforo. Para esta aplicación, se eligió optimizar una metodología previamente desarrollada y validada para: aminoácidos en orina y cerebro de ratón. Con las condiciones finales optimizadas utilizando este método se lograron detectar 14 aminoácidos en plasma, todos resueltos de sus formas enantioméricas (L- y D-). El método fue completamente validado y aplicado para la cuantificación del perfil de aminoácidos en plasma de pacientes con BD. Todos los valores obtenidos de la cuantificación estuvieron dentro del rango de los valores previamente publicados.

En resumen la CE-LIF es una técnica de alta resolución, sensible y selectiva, de alta velocidad que posee las mismas ventajas que se mencionan para la CE con detector UV, pero con la ventaja adicional que aporta el detector LIF en el aumento de la sensibilidad y selectividad. Dentro de los mayores inconvenientes de esta técnica se puede mencionar justamente el proceso de derivatización, ya que no todas los compuestos pueden pasar por este proceso, el cual se encuentra además condicionado por el tipo de láseres comerciales disponibles y el limitado número de longitudes de onda. Además entre otros inconvenientes, durante el proceso de derivatización muchos de los reactivos utilizados usan compuestos tóxicos como el cianuro de sodio, que suelen generar un largo número de subproductos del mismo compuesto.

Una de las ventajas de este método sobre los que se encuentran publicados es la posibilidad de separar y medir los aminoácidos quirales. La importancia de medir las formas específicas de estos está creciendo en el área de la neurociencia donde aparentemente las formas D- de los aminoácidos parecen tener una función como neuromoduladores y probablemente tengan un papel considerable en el desarrollo del sistema nervioso. Más aún para la enfermedad de BD, se ha publicado que el aminoácido D-serina

pudiera ser un biomarcador para psicosis, el cual además se ha demostrado que produce un efecto antidepresivo en roedores y humanos cuando se ha usado como un tratamiento médico. Sin embargo el papel de la D-serina en pacientes con BD que son tratados con una infusión de ketamina no se había establecido antes. Aunque los datos no fueron publicados se encontró que independientemente de la co-medicación utilizada (litio y ácido valproico) los niveles de D-serina fueron mayores en el grupo de pacientes que no responden al tratamiento ($p= 0.038$). Por lo tanto, la cuantificación de D-serina en el plasma representa un potencial biomarcador de la respuesta al tratamiento con infusión de ketamina. Finalmente, este mismo resultado fue validado en un grupo nuevo de pacientes con trastorno depresivo mayor (MDD) tratados con infusión de ketamina, donde los misma tendencia fue observada (datos no mostrados).

Finalmente el Capítulo 7, muestra la comparación de los resultados de todas las técnicas englobadas una discusión general basada en parte por los aspectos generales que ya se encuentran citados en la literatura y complementados por los distintos desafíos y problemas enfrentados durante el desarrollo de las diferentes publicaciones.

Como resumen final, cada plataforma analítica presenta características propias que las han hecho adecuadas para un grupo de metabolitos. En este sentido cada técnica analítica caracteriza una parte distinta del metaboloma. El metabolismo formado por una red de conexiones compleja en la que participan una vasta cantidad de enzimas, transportadores y mecanismos genera por tanto una cantidad muy grande de metabolitos en concentraciones muy diversas en el organismo. Por lo que cada una de las técnicas analíticas es capaz de ver un grupo distinto de metabolitos según las propiedades físico-químicas de los metabolitos y el límite de sensibilidad de la técnica (RMN, CE–UV<MS). Sin embargo, incluso usando cualquier plataforma acoplada a MS seguirán presentes algunos compuestos que estén por debajo de su límite de detección. Durante la realización de este trabajo doctoral se demostró que GC–MS, RMN, CE–UV y LC- MS son técnicas que pueden trabajar de forma fiable para obtener el perfil del análisis no-dirigido en muestras biológicas. Por otro lado, cuando el análisis es dirigido es necesario desarrollar un método de análisis para los compuestos de interés. Aquí, una de las herramientas presentadas para cumplir con ese deber fue la CE–LIF. Esta técnica representa una buena opción debido a su alta sensibilidad y especificidad. Todas estas plataformas analíticas forman parte de las herramientas para obtener un perfil integrado más cercano a captar el metaboloma.

Por último, el campo de la metabolómica está en continua evolución y con ello, las plataformas de análisis y herramientas informáticas y de tratamiento de datos. Con esta idea en mente, es posible que nuevas pruebas clínicas de diagnóstico y pronóstico estén cerca de realizarse a través de uno de los enfoques de la metabolómica utilizando cualquiera de las técnicas de alto rendimiento.

1. Introduction

1.1. The evolution and development of *–omics* approaches in systems biology

Years ago, the thought of in–depth knowledge of biomolecular processes in a disease progression, and the outcome of novel diagnostic tests that could lead to personalized health care were unimaginable. It was after the realization of human genome mapping that the research field was triggered by the interest in approaches that could relate gene expression to phenotype[1]. Then new disciplines emerged trying to answer this using interrogative measuring gene expression (transcriptomics) and protein translation changes (proteomics)[2]. Billions of dollars and intensive work hours were pumped into a huge industry built on the new *–omics* sciences, finding not enough information for an understanding of the phenotype in the living systems[2]. Then, and still partially now, genetic discovery had exceeded the ability to understand completely gene function, and one of the bottlenecks that existed was the need for high–throughput tools to evaluate the gap to phenotype[3–4].

In this sense, it was with the change in the central dogma of molecular biology that the revolution of systems biology started. Scientists realized that mapping genomes was not the universal solution needed to understand complex biological questions. No longer true was the central dogma that had established the unidirectional flow of information from gene-to-transcript-to-protein-to-metabolite; now, it is well known that cellular processes are in fact intimately networked with many feedback–loops among the different levels[5]. In addition, biochemical pathways were not seen as isolated systems; their interconnections pictured metabolism as dynamic and complex networks. Here, metabolites were regarded as real biological endpoints of metabolism, and thus capable of characterizing the phenotype in an organism[6]. Based on the paramount importance of metabolites, Oliver, S. G. was the first who introduced the term "metabolome" in 1998[7], describing the total low molecular weight or metabolite content of an organism. Later the expression "metabonomics" was coined by Nicholson, J. K. in 1999[8], where it was defined as "*the quantitative measurement of the dynamic multi–parametric metabolic response of living systems to pathophysiological stimuli or genetic modification*". A little time later, the expression "metabolomics" was introduced by Fiehn, O. in 2000[9]. This discipline was defined as "*the unbiased global survey of all the low–molecular–weight molecules or metabolites in a biofluid, cell,*

tissue, organ, or organism[10],[1, 8, 11]. Now, the words metabolomics and metabonomics are often used interchangeably or equally. The high value of this emerging approach was based on work that started to link metabolic changes with genotype gap.

Along with the different *–omics* approaches, all constitute the contemporary systems biology framework focusing to extract useful information, including: the analysis of genetic variation (genomics); gene expression (transcriptomics); gene products (proteomics); and their metabolic effects (metabolomics)[12–13]. Figure 1.1 shows the representation of the traditional central dogma (A), versus the dynamic metabolic status (B); and the past visualization of linear pathway (C), versus the complex interacting metabolic network (D).

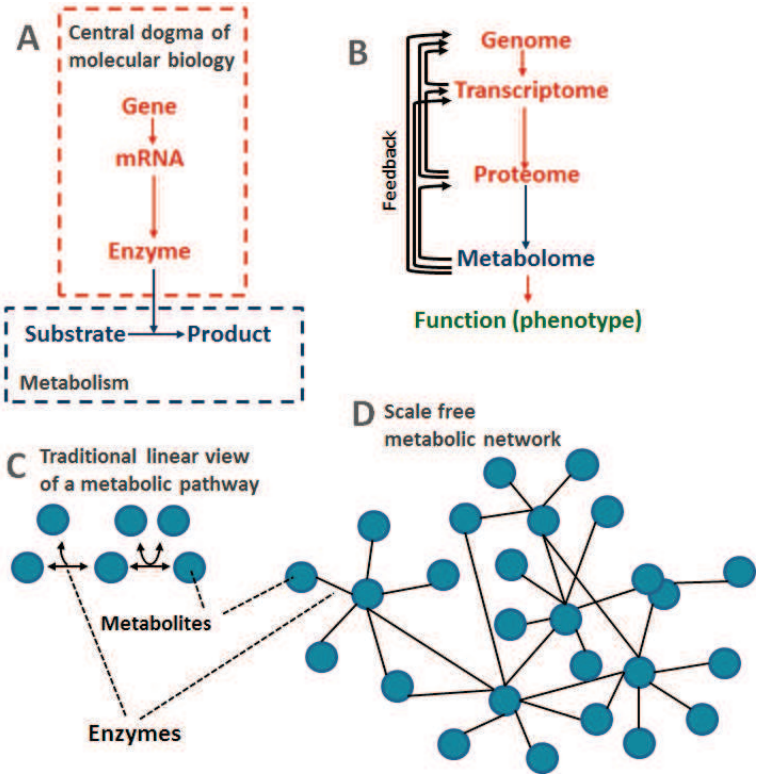


Figure 1.1. Change of scheme of systems biology. A) Traditional central dogma. B) General schematic of the *–omic* organization. (C) Traditional linear view of a metabolic pathway and D) the now accepted view of scale–free connections[5]. *Key*. Metabolome is the complete collection of metabolites in an organism (same meaning for genes–genome, mRNA–transcriptome and proteins–proteome).

Following this lead Fiehn, O., in 2001 defined that transcriptomics characterizes the population of mRNA species in a cell at a specific time and set of conditions; proteomics addressed the challenging problem of defining changes in protein expression, protein dynamics and post–translational modifications; whilst the emerging field of metabolomics measured changes in populations of low–molecular–weight metabolites under a given set of conditions[10].

In this post–genomic era, the wealth of knowledge in the *–omics* field has allowed researchers to now perform data integration from different platforms, offering new ways for determining causes of human diseases and in finding possible cures. Additionally, the *–omics* data should give new possibilities for the drug discovery process and toxicology screening[14]. Nevertheless, current bioinformatics tools cannot easily incorporate multi *–omics* platform data[15,] owing to most of the diseases being complex and multi–factorial, and that the disease phenotype is determined by the alterations of multiple genes, pathways, proteins and metabolites (at cellular, tissue, and organism levels)[16–17]. Also, the difficulty in this task is increased by the fact that each process at each level occurs at different points in time.

The application of metabolomics as a detection tool was first developed in the pharmaceutical industry for toxicity predictions and screening. To date clinical diagnosis, prognosis and treatment selection are increasingly dependent on the use of molecular tools, predictive modelling and deep biological information (e.g. genetic, metabolic or physiological data)[18]. For example the actual management of patients with cancer is well suited to a personalized approach where new diagnostic tumour classification and prognostication has been achieved through the use of genomic and proteomic sequencing[19]. In addition, relatively a new approach using metabolomics has just begun to enter the mainstream of cancer diagnostics and therapeutics[20–21]. The development of powerful and effective pharmacologic targets has also highlighted the necessity for individualizing drug therapy to select those patients most likely to respond to treatment, to minimize the occurrence of adverse drug reactions, and to maximize the desired therapeutic effect. Finally metabolomics will be a major research area in which to develop dynamic system–monitoring tools that lead to diagnosis. Static, single time–point biomarkers will be replaced by dynamic biomarkers that can identify whether the system is moving towards or away from health[16, 22].

1.2. Metabolomics

The word "metabolism" comes from the Greek word "μεταβολή" (metabolē), which means *change*. Metabolomics offers a practical approach to measuring the metabolic end points (metabolites) and their *changes* that link host genetic and environmental factors with phenotype. Thus, metabolites are the result of the combination of genetic and environmental factors such as diet, age, ethnicity, lifestyle and gut microflora[2]. For example, in a recent study about the susceptibility of metabolic phenotypes to be modulated by diet, authors were able to distinguish meat-eating individuals from the vegetarian group, due to alterations of bacterial composition in response to diet[23].

When any level of physiology in an organism is altered, e.g. the levels of proteins in response to environmental or pathological cues, the organism will attempt to maintain homeostasis in the face of challenge. Complex organisms such as human beings possess multiple mechanisms that provide inherent flexibility to manage perturbations of physiology and metabolism. In an effective response to this, the system is kept under normal regulation called allostasis[22]. This flexibility is manifested in "biological robustness", which characterizes the propensity of a living system to maintain effective functioning in the face of internal and external stressors. But when the system is no longer able to respond effectively, a disease stage will be developed, and as a consequence metabolic pathways will be disrupted, generating up- and down-stream of a set of metabolites (Figure 1.2). As an outcome, these changes are amplified in the metabolome to give an accumulated and highly sensitive description of the physiological state[24]. Metabolomics offers potential advantages that classical diagnostic approaches do not, based on following the discovery of a suite of clinically relevant biomarkers that are simultaneously affected by the disease[25–26].

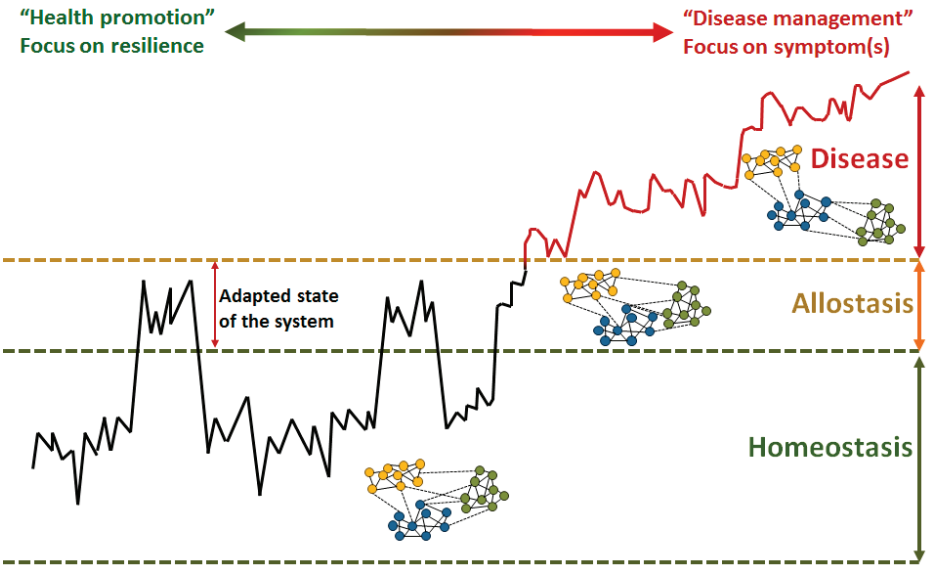


Figure 1.2. Illustration of the dynamic system regulation (taken from van der Greef, J., 2013[22]). An organism that could respond effectively to any alteration (environmental or genetic) is kept in the "allostasis" area and herein considered as "healthy". When the organism no longer responds correctly, a disease state is developed.

The measurement of all metabolites in an organism, and even in a sample, is far from possible, for two main reasons. First, the quantity of metabolites is very large, e.g. for the plant kingdom alone there is an estimated 200,000 metabolites, and for the human this figure is expected to be even higher[27–28]. Second, there is a high diversity of metabolites present in a wide range of concentrations. These are generally labile compounds with different chemical properties[5], including wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility, charge) properties. The degree of diversity includes ionic polar compounds, such as oxalate, low molecular weight (MW), volatile organic metabolites (urea), to the higher molecular weight, polar (carbohydrates) and non–polar (terpenoids and lipids) metabolites. In addition, concentrations also extend over an estimated 7–9 orders of magnitude in concentration (pmol–mmol)[29]. Moreover, their stereochemistry leads to different biological functions[30].

In this context there is no an actual analytical strategy that could cover this challenge. Nevertheless, the simultaneous determination of a high number of metabolites in a sample that could characterize the

physiological stage of an organism is possible. For the study and analysis of metabolites, the metabolomic field has tackled several strategies that are still evolving[27, 31]:

1. *Target analysis.* This precise and quantitative approach is intended to analyse one or few metabolites generally metabolic-pathway related (typically less than twenty[32]). This approach could be driven from known biology, potentially derived from a previous discovery study (e.g. validation of potential biomarkers) or when low limits of detection are observed e.g. substrates and products of enzymatic reactions[27, 33]. In addition, targeted strategy is also used as part of other approaches, such as non–targeted analysis, commonly for metabolite identification, e.g. by tandem mass spectrometry (MS/MS) of predefined metabolite signals. This strategy has been applied in traditional analytical chemistry and biochemical assays in clinical laboratories.
2. *Metabolic profiling.* Initially referred to and focused on a specific group of metabolites, e.g. amino acids, free fatty acids, etc.[5]. Today, this term is a synonym for non–target analysis.
3. *Non–target analysis.* Is the unbiased and simultaneous determination of multiple metabolites in a single run: enough to represent the metabolic status of an organism (metabolites here includes commonly unknown compounds)[33–34].
4. *Metabolic footprint.* Also called exo–metabolome or secretome. The aim of this approach is to measure the extracellular metabolites, e.g. the medium components in a cell culture. The footprint of a cell culture shows the medium components, less substrate uptake, plus metabolites secreted by cells[5].

For the approaches mentioned above, non–targeted analysis is the most widely used in metabolomic studies. Initially in this approach, there is no particular interest in identifying any compound *a priori,* but to compare patterns of metabolites that change between the study groups[35]. In this sense, as metabolism integrates information from gene expression, the microbial gut and a wealth of environmental cues. Therefore, each organism will exhibit a distinct and particular metabolic pattern that later could possibly confound results[11]. Regardless of the variability in the metabolic profile, some characteristic metabolites will be specifically up/down regulated in response to a particular disease, toxin exposure, or environmental or genetic alteration[33, 36]. These differences are normally found through multivariate and chemometric tools. The non–targeted strategy potentially offers *de novo* target discovery, where the

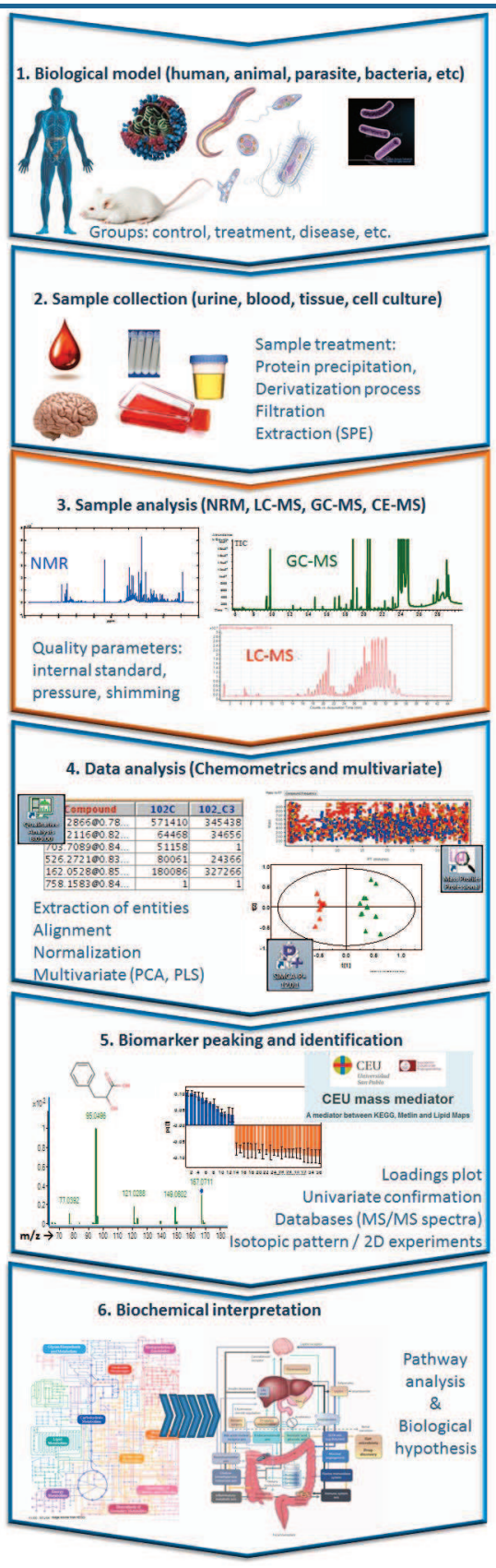
exploration of the chemical space is limited only by the sample preparation and the characteristics of the analytical technique (sensitivity and coverage)[37].

Along with the evolving approaches, the field of metabolomics has already produced a vast array of specialist terms which are also still evolving (see the glossary section), e.g. *lipidomics* was defined by the characterisation of chemically distinct lipid species in cells and the molecular mechanisms through which they facilitate cellular function[38].

At this point, it is important to keep in mind that the study of the dynamic metabolism in the human being is abruptly complex; whilst the human genome sequence has been completed, the vast entourage of commensal microorganisms, estimated to be more than 1,000 different species that live in symbiosis, have not[39]. Lindon, J. C. and Nicholson, J. K. in 2008 reported: "Animals, including humans, can be considered "*superorganisms*" with an internal ecosystem of diverse symbiotic gut microflora (often with unknown genomes and functional ecologies) whose metabolic processes interact with the host"[40]. Here, metabolites originated from the interactions with symbiotic organisms such as gut microorganisms are regarded as "*co–metabolites*"[39]. Importantly, the gut flora can modify the host metabolome through co–metabolism of compounds from dietary intake such as phenols and phenylalanine[11]. The gut dysbiosis contributes and influences the mammalian processes; Nicholson, J. K. in 2005[39] has shown the influence of the microbiome in the progression of prevalence human diseases such as diabetes, obesity, cancer, non–alcoholic fatty liver, inflammatory bowel and cardiovascular diseases[36, 41].

1.3. Non–target analysis workflow

There are a large amount of examples that could show the development and the application of non–targeted analysis in systems biology for toxicological screening, diagnosis biomarkers, prognosis of disease and drug efficacy, amongst many other applications. In this thesis some examples are shown. The workflow normally followed by most of metabolomics studies in non–target analysis is shown in Figure 1.3.



The workflow in metabolomics (Figure 1.3) starts setting the biological model (1), where according to the objectives, economical budget and resources (e.g. analytical technique, laboratory material, physicians, etc.), the number of groups and the size within each group, the metabolomics project is established. Then the sample is collected (2) following some standard procedures in order to further treatment based on the technique of analysis. After samples are run using a specific analytical platform (3), the data acquired is treated by multivariate and chemometrics tools. Potential biomarkers in the study are statistically peaked up and identified (4). Finally, their dis–regulation is explained by the biochemical pathways and, if possible, a biological or molecular hypothesis is performed (5). These biomarkers are normally derived from key sources of phenotypic variation, e.g. dietary, gut microbiomic, xenobiotic (e.g., drug use), host genetic influence, etc. This workflow is largely explained in the following sections.

Figure 1.3. Workflow followed by metabolic non–target analysis. Note: Step 3 shows an example of non–target profile in GC–MS, LC–MS, and ^{1}H–NMR.

1.3.1. Biological model

Extensive metabolomics studies have been done in human biological models, nevertheless studies in animals, parasites, bacteria, and virus are also acceptable candidates in this field. Here the selection of appropriate groups in a biological model is not always an easy task. The objective is to plan and conduct experiments in order to extract the maximum amount of information with the fewest number of experimental runs. For an animal model there is a reasonable concern of working with the fewest amount of animals, which has as a drawback that the size of the groups is always small. Although, using inbred strains with homogenous genetic background, age distribution, controlled food and water supply, and defined living conditions (caging, standardized day–night cycles), enhances the readability and reduces bias in the study. In the human model, one of its overarching points is to highlight the importance of selecting the appropriate comparison groups. Ideally, the comparing group must be represented by the exact characteristics of the study group marked only by the absence of the disease, treatment or characteristic that is the objective of the study. Many studies had failed because the comparing group was badly chosen. For example, in a study that is being carried out on hospitalized patients, the comparing group (even if they are "*healthy*" persons) should be hospitalized too. Otherwise, medications and their influence in the metabolism will show up as good biomarkers, distracting from the important pathways underneath. In addition, the selection of patient gender, age, ethnicity and clinical history are aspects that always need to be considered and should be as homogeneous as possible.

1.3.2. Sample collection

Biological fluids are rich in a wide range of molecules, such as proteins, lipids, and small molecular weight metabolites[42]. Sample collection is primarily driven by objective, the experimental design and sample availability[33]. Originally metabolomics studies were conducted basically in biofluids, many of which were usually easily obtained, such as urine and plasma, both suitable representatives of the general metabolism of an organism and practical to use in the clinic[35]. While a plasma sample better represents the real physiological stage where many different types of intermolecular interactions are occurring simultaneously (defined as the 'snapshot' of the global ''*metabolome*" at the time of

collection[42]), urine samples represent the time–averaged end products of metabolism, which captures the excretion of exogenous compounds coming from diet, drugs, etc.

As time progresses, the study of other biological fluids (e.g. saliva, cerebrospinal fluid, seminal fluid, amniotic fluid, breast milk, etc.) with the investigation in cell culture and tissue (either from biopsy or post–mortem) are becoming more frequent[2, 43]. Although, protocols for the collection, storage, and processing of these kinds of specimens are less established than those for blood and urine.

For sample collection all aspects should be considered, the group involved (physicians, surgeons, pathologists, and technicians) should be aware that trivial mistakes can cause the whole experiment to fail. The reproducibility and reliability of results in any research is threatened by different sources of bias. Clinical bias may include differences in the sample collection protocol: type of tube; amount of time on ice; time of sampling (diurnal control); type of anticoagulant; the same type of blood sample (serum and plasma are not the same); storage conditions; freeze–thaw cycles; length of storage time; freezer temperature[31, 44]. These aspects are highly relevant when sampling is collected at different sites (laboratory, hospital or centre) and transported to a central laboratory for processing. Also, aspects such as the randomization of the order of preparation and analysis are important to avoid analytical bias. A good article discussing sample preparation, trends and challenges used in metabolomics specifically in LC–MS was reviewed by Vuckovic, D., 2012[45].

If possible, the standardization and control of sample collection procedure, diet, exercise and medication can have a high impact on the reliability of the metabolomic test result. It is essential to understand, minimize, and control sources of both analytical and biochemical variation. For example, Teahan, O., in 2006[44] showed the importance and the effect on sample collection protocol on NMR–derived metabolic profiles of plasma and serum. Variations in the serum clot contact time, storage conditions prior to centrifugation, freeze–thaw cycling, sample collection onto and in absence of ice, and sample stability at room temperature were tested attempting to establish a general protocol for use in metabolomics[44]. For experimental animal modelling, Bollard, M., 2005[46] explored the differences in the metabolic pattern under various normal physiological conditions in animals: gender, age, diet, species, strain, hormonal status, diurnal variation, and stress by urinary ^{1}H NMR data[46]. Furthermore, Bictash, M., 2010[42], showed an overview of the factors that affect the human metabolic profile: gender, ethnicity,

population characterization, diurnal variation, age, body mass index (BMI), and diet were described in large–scale metabolic phenotyping. It is important that after multivariate analysis the metabolite changes have no false biomarkers, resulting e.g. from sample preparation.

The choice of sample preparation method affects both the observed metabolite profile and the data quality[45]. The treatment applied to a particular type of sample should be specific according to the objective of the study and the analytical platform that will be employed. For example, in the case of using gas chromatography (GC), the biological sample normally should go through derivatization as the sample must be volatile. In addition, some aspects should always be considered before the analysis, such as dilution of the sample in the case of urine, amount of solvent when the metabolites are going to be extracted from any tissue, filtration of sample in case any precipitation is observed, sample manipulation and care (e.g. sample should be thawed on ice), etc. Much more care should be taken with sample manipulation in the case of tissue in order to avoid cross contamination and to extract metabolites without bias. For metabolites from cells or tissues, a concern is the speed of quenching of metabolism. Without a deep understanding of the capabilities and limitations of the sample treatment method used in a study, the accuracy of biological interpretation of collected data may be compromised.

1.3.3. Sample analysis

Once the sample is properly collected and treated accordingly to the analytical platform that is going to be used, the next step is the processing of sample. The availability of valid data is a fundamental prerequisite for a successful outcome of any study. Separation science plays a critical and often overlooked role in metabolomic studies. As previously commented, the analytical platforms have a paramount role in metabolomic studies: their characteristics and applications are the principal objective in this work and therefore, they are going to be described more extensively in the course of the chapters of this work.

1.3.4. Analytical platforms used in metabolomics

It was indeed with the development of robust and high throughput analytical techniques, along with the realization of chemometric and pattern recognition analysis that the research in metabolomics took off and became accelerated. With techniques such as nuclear magnetic resonance spectroscopy (NMR) used in the beginning, and later, mass spectrometry (MS), the simultaneous measurement of large number of metabolites on a single biological sample was possible. The principal reason that NMR and MS are still the main techniques used and exploited in this field is that both are capable of detecting a significant number of metabolites and, more importantly, both give specific molecular structural information, and thus afford the possibility of identification. Therefore, NMR and MS have become established and robust tools for metabolite profiling, providing fast and reliable data[9].

In a review paper by Griffiths, W. J., in 2010[37] the authors showed the number of publications indexed in PubMed for MS and NMR–based techniques from 1999 to 2009 (Figure 1.4A). Here results showed that from 2005, MS–based metabolomic studies were preferred over NMR (at a ratio of 2:1). In addition, in a more recent review work by Kuehnbaum, N., 2013[47] (Figure 1.4B) the number of publications in metabolomics via Web of Knowledge over five years (2008–2012) depicted that an average of 46% of articles used NMR. Even though the search parameters and the databases used were not the same; Figure 1.4A depicts the increasing trend to use MS over the time, while Figure 1.4B shows the high rate that NMR is still applied in metabolomics (nearly 50%) and that among the hyphenated techniques coupled to MS the most used was liquid chromatography (LC).

The analytical platform selection is one of the main questions when a metabolomic study is being planned. The answer will rely on the purpose of the study, the type of sample, and the available of resources. For a global profile, employing more than one technique (multiplatform) or separation mode is supposed to be a good option, based on the fact that each technique will contribute to gaining a broad perspective of the metabolome[48]. The aim of these procedures is to produce biochemically based non-targeted profiles that are of diagnostic or other classification value. For targeted metabolomics the decision depends on choosing a well suited technique for the group of metabolites of interest. For example, gas chromatography coupled to MS (GC–MS) is well suited for the analysis of stable and volatile metabolites such as fatty acids, while LC–MS has proved to be appropriate for lipids. The choice

of which technique to use will depend on volume sample, sample preparation requirements and instrument availability.

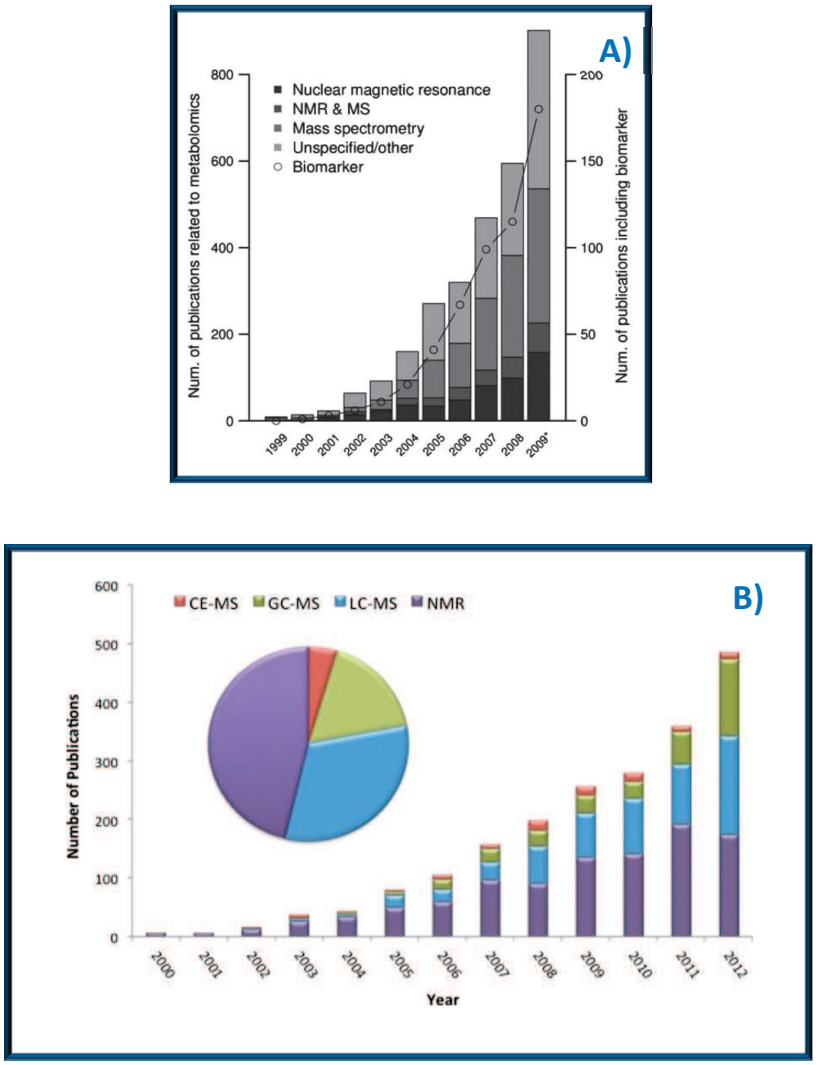


Figure 1.4. A) Number of papers in PubMed covering search terms related to metabolomics, as compared with the set of techniques: "nuclear magnetic resonance/NMR", "mass spectrometry/MS" and "biological marker/biomarker" (figure taken from Griffiths, W. J., 2010[37]). B) Literature survey of published research articles in metabolomics via Web of Knowledge 5.8, excluding reviews and abstract proceedings (taken from Kuehnbaum, N. L., 2013[47]).

The first works involving the concept of metabolomics were based on proton nuclear magnetic resonance (^{1}H NMR) on biofluids, cells and tissues[49]. Subsequently, this concept was extended on the application of GC–MS and LC–MS[9]. Currently, the range of platforms widely used for metabolomic analysis, include ^{1}H NMR spectroscopy, direct infusion mass spectrometry (DIMS), and MS combined with a separation technique, e.g. GC–MS, LC–MS and capillary electrophoresis coupled to MS (CE–MS)[50]. Other analytical techniques have been used in metabolomics studies and these are listed in Table 1.1. For CE–UV technique, this will be discussed in depth in one of the chapters of this work. The main limitation of using these techniques (FT–IR, LC–EC and CE–UV) later were the low level of structural information provided for biomarker identification that can be achieved. This made them less used in metabolomics.

Table 1.1. Analytical techniques applied in the time course of metabolomics.

Abreviation	Technique	Ref
^{1}H NMR	Proton nuclear magnetic resonance	Beckonert, O., 2007[52].
(HR–MAS) NMR	High–resolution magic–angle–spinning NMR	Beckonert, O., 2010[53].
LC–MS	Liquid chromatography–mass spectrometry	Zhou, B., 2012[54].
UPLC–MS	Ultra–performance liquid chromatography–MS	Want, E. J., 2010[55].
LC–EC	LC using electrochemical array	Gamache, P. H., 2004[51].
LC–NMR	LC coupled to NMR	Wolfender, J. L., 2003[56].
HILIC–UPLC–MS	Hydrophilic interaction–UPLC	Spagou, K., 2011[57].
LC x LC–MS	2 dimensional LC coupled to MS	Guo, K., 2011[58–59]
GC–MS	Gas chromatography–mass spectrometry	Garcia, A., 2011[60].
GC x GC–MS	2 dimensional GC coupled to MS	Mal, M., 2012[61].
FT–IR	Fourier transform infrared spectroscopy	Ellis, D. I., 2006[62].
DIMS	Direct infusion ESI MS	Han, X., 2003[38].
MALDI–MS	Matrix–assisted laser desorption–MS	Milne, S. B., 2013[63]; Vaidyanathan, S., 2005[64].
FT–ICR–MS	Fourier transform ion cyclotron MS	Aharoni, A., 2002[65].
CE–MS	Capillary electrophoresis–MS	Ramautar, R., 2009[66]; Barbas, C., 2011[29].
CE–UV	CE–ultraviolet detector	Barbas, C., 2011[29].

Adapted from Hollywood, K., 2006[5].

Independently of the analytical technique employed, one of the cornerstones of any metabolomics study is the acquisition of high–quality data. The assessment of the instrument stability and quality of the data acquired is of paramount concern, e.g. for MS–based experiments the use "quality control" samples, internal standard and/or standard mix is a part of the run routine.

This present work shows most of the analytical platforms applied in metabolomics. It aims to provide a comprehensive overview of the analytical techniques, and therefore at the beginning of each chapter a resume of the basis, advantages and inconveniences of each technique is presented. In addition, the different challenges faced with each study are pointed out and discussed at the end of each resume. Additional aspects such as the control or monitoring the acquisition of the data, common troubleshooting, and importance of sample preparation are also commented on.

This thesis work has been divided in three sections: first section shows an example of *analytical method development and validation* using two different analytical platforms (GC–MS and LC–MS). In this section the analytical methodology of Chapter 2 shows the optimization of an extraction procedure focused on the elaboration of a non–target profile for breast milk and characterized by LC–QTOF–MS and GC–Q–MS.

The second section exhibits the metabolomic applications using *non–targeted analysis*. Here Chapter 3 summarizes and highlights the use of ^{1}H–NMR. In addition, Chapter 4 shows the employment of CE–UV as a complement of ^{1}H–NMR, proving the advantage of multi–platform coverage. And in Chapter 5, the application of LC–QTOF–MS demonstrates its high capabilities and contribution in the metabolomics field. The third section shows an example of *target analysis*; the optimization of an analytical method for the quantification of amino acids in plasma by CE–LIF (Chapter 6). Finally an extensive discussion confronting all these techniques is presented in Chapter 7. A more extensive description of these chapters can be read below:

Chapter 2 describes the development and validation of a non–target profile for breast milk. The aim of this project was to develop a common extraction procedure of metabolites from breast milk which could allow for its characterization by two analytical techniques (LC–MS and GC–MS), therefore

expanding the metabolite coverage. The final purpose was its later application in metabolomics studies, e.g. in the study of the changes in composition of breast milk according to time.

Chapter 3 is the characterization of the global metabolic phenotyping of patients with acute pancreatitis in the emergency unit. Here the metabolic characterization is done using ^{1}H NMR from urine and blood samples and the aim of this study was to determine the diagnostic potential of a metabolomic approach in the diagnosis and prognostic staging of acute pancreatitis and to ascertain the clinical utility of this tool in the analysis of acute abdominal pain in the emergency setting.

Chapter 4 is the urinary metabolic phenotyping of slc26a6 (chloride–oxalate exchanger) null mouse model. This study shows the application of CE–UV in the field and moreover the combination with NMR spectroscopy, providing the increased coverage of the urinary metabolome and the use of multi–platforms in metabolomics. The aim of this study was to show the direct relationship between urinary oxalate and a range of urinary metabolites and to show clear gender–dependent slc26a6–related differences in metabolism.

Chapter 5 is the description of plasma metabolic patterns of resistant–treatment patients receiving ketamine for bipolar depression. Here, a comparison between responders and non–responders patterns to the treatment by LC–QTOF–MS is performed. In this pilot study, the purpose was to identify potential biomarkers for the response and non–response to the treatment and their associations with the pathway disruption.

The Chapter 6 in relation with the previous one shows the target analysis of amino acids in plasma of resistant–treatment patients with bipolar disorder. Here it is well known that amino acids in bipolar depression are altered compared to healthy persons[67–68], but no study has been done in relation to ketamine treatment. This is why the first step to establish any relationship with the response to ketamine is to develop a target approach for the acquisition of amino acid profile along with method validation by a very selective technique such as CE–LIF.

Although these analytical techniques are going to be extensively discussed at the beginning of each chapter; a brief and general description is shown next.

1.3.4.1. Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR is a very commonly used analytical tool that has been used in metabolomics from the beginning and through the development of this field. This technique is based on detecting the specific resonance absorption profile by a specific nucleus when this is placed in a magnetic field (generally from 600 to 750 MHz). This nucleus, normally contained in the metabolites, will absorb depending on the strength of the magnetic field, the type of the nucleus and the nature of the chemical bonding (chemical structure). This makes NMR a non–selective technique, almost considered as universal. It requires little or no sample preparation, is cost-effective, unbiased, rapid, robust and reproducible, quantitative, and non-destructive, thus making it highly suitable for non–target analysis. The principal disadvantage of this technique is its low sensitivity compared to other techniques, such as MS. Importantly, ^{1}H NMR spectroscopy has been considered the analytical workhorse for metabolomic research[69]. In addition, NMR is useful for metabolite profiling of intact biofluids and more recently intact tissues have been analysed using magic–angle spinning spectroscopy (HR-MAS-NMR)[53]. Moreover, a promising option in the analysis of complex mixtures that technological developments have allowed is the direct parallel coupling of LC systems to NMR for biofluids, although this is still under optimization[70]. Good reviews of NMR have been published elsewhere[42, 69].

1.3.4.2. Mass Spectrometry (MS) coupled to a separation technique (LC, GC or CE)

A hyphenated technique is developed from the coupling of a separation technique and an on–line spectroscopic detection technology, i.e. LC–MS, GC–MS, CE–MS, LC–NMR, LC–FT–IR, etc. In recent years, hyphenated techniques linked to MS have received increased attention because they combine the strength of a separation technology (CE, GC, and LC), and the qualitative (and quantitative) information leading to the identification of unknown compounds in complex samples (MS). In the history of the improvement of MS, the successful coupling of GC–MS during the sixties led to the idea of coupling LC to MS and finally of using CE. It was with the development of new interfaces that managed to revolutionize the analytical scenario, and many MS applications in different fields were created and used

as routine[30]. The application of these strategies improved some of the limitations of linking direct infusion to MS (DIMS), such as the detection of isomers, isobars and the suppression effect caused by molecules competing for ionization, giving the possibility of separating and concentrating different classes of compounds according to their physical–chemical properties. Nevertheless, DIMS is still contributing to the metabolomics field, with applications mainly in the microbial and plant fields[27]. New applications e.g. through shotgun lipidomics first described by Han and Gross[38, 71,] have been explored, where extracted lipids are analysed by direct infusion in ESI–MS or ESI–MS/MS. Generally high mass resolution analysers such as QTOF with 5–ppm mass accuracy and FT–ICR with less than 1 ppm mass error are employed[27]. Full details on the methodology, advantages and drawbacks are outside the scope of this work and are discussed elsewhere[72–73, 37].

The hyphenated systems present wide versatility because each one can be worked in different ways or modes, i.e. LC–MS can use different type of buffers, columns, gradients and temperatures making it suitable for a high number of applications. Each separation technique has different requirements, advantages and drawbacks. GC is an excellent separation technique which, through the use of fused–silica columns several meters long, achieves high peak resolution, efficiency and reproducibility. However, GC is limited to thermally stable, volatile small compounds, or compounds that can be made chemically volatile (e.g., by trimethylsilylation). Separation of compounds is determined by volatility of sample components. For non–targeted and targeted analysis, GC has been coupled to FID or MS detectors[27, 74].

LC is a powerful technique suitable to be used to resolve different types of groups, from hydrophilic to non–polar. The separation of each group of compounds is dictated by the properties of the solutes when interacting between the stationary phase from the column and the mobile phase passing through it. LC has different chromatographic modes that can be tailored to the separation of specific classes of compounds. This includes: reversed phase (RP), normal phase (NP), ion exchange, chiral, size exclusion, hydrophilic interaction chromatography (HILIC), and mixed modes. Here, RP–LC is an extensive mode used to separate medium to non–polar compounds, while HILIC–LC is used for the separation of polar metabolites. For detection systems, UV/Vis absorption and MS has been applied, although for metabolomics only MS has been kept for extensive use. Now with the introduction of ultrahigh–pressure LC (UHPLC), instruments with columns packed with sub–2µm particles are capable

of withstanding pressures up to 12,000psi. Chromatographic results show sharper peaks (higher resolution), sensitivity and an exceptional increase in the speed of analysis[74].

Although CE is a fast and high separation technique, this was the last one to be used within the metabolomics field. In any electrophoretic method, charged molecules are moved through a liquid or a gel by an electric force. Here, the conduit for the molecule electrophoretic mobility is a capillary in which the molecules are separated based on their size to charge ratio. CE has different separation modes, among them capillary zone electrophoresis (CZE) and micellar electrokinetic chromatography (MECK) are the most often applied. The use of various detection modes (e.g., ultraviolet detector (UV), laser induced fluorescence (LIF), conductivity, amperometry and MS) further enhance the capabilities and applications of CE. In general, the advantages of CE include high peak efficiency, short analysis time, and the requirement of small sample volumes. Its selectivity is the main reason to use it for the analysis of small and polar organic ions. However, the main drawback for a metabolomics study in CE is its low reproducibility of migration time. The additional advantage of CE coupled to MS is that almost any charged species can be infused into MS. Although among the typical constrains of MS, buffer systems and separation modes are restricted to volatile buffers such as acetate and formiate[29, 75–76].

Finally considering only MS, only those compounds that can be ionized can be detected, this by addition or removal of a proton or another ionic species or by fragmentation acquiring charge. Here MS has numerous configurations for optimizing and improving metabolite analysis. MS has a wide variety of instrumental equipment based on the interface, the ionization mode, mass analyser, and detector to be part of. Advances in MS instrumentation, particularly in interfaces and scanning speed, allowed the detection of small as well as large molecules in complex mixtures at sensitive levels[48]. For ionization interfaces, electro spray ionization (ESI), atmospheric pressure photo ionization (APPI), and atmospheric pressure chemical ionization (APCI) are the interfaces used in LC–MS. Although for metabolomics applications: ESI is the most common interface for LC and CE, while electron impact ionization (EI) is widely used for GC–MS[30]. Moreover, matrix–assisted laser desorption ionization (MALDI) is now employed in many laboratories worldwide for spatially resolving proteins, peptides, lipids and small molecules directly from tissue sections[77].

A mass analyser such as quadrupole time of flight–MS (QTOF) is one of the most used for metabolomics investigations coupled to LC. This is because QTOF is capable of providing structural information for compound identification through MS/MS experiments. The quadrupole (Q) may act as selecting, isolating or as a fragmentation cell of specific compounds, while the TOF provides high mass resolution and accuracy. Other mass analysers such as Fourier transform ion cyclotron resonance (FT–ICR), ion trap (IT) and triple quadrupole (QqQ) can generate fragmentation of compounds (MS/MS data) that can be used for the identification of compounds. Each mass analyser has different mass resolution, accuracy, and sensitivity, among other characteristics. These are discussed in depth elsewhere[78]. For example, FT–ICR offers the highest resolving power and accuracy of all MS instruments to date, while QqQ is used for high target throughput and sensitive analysis. Specific characteristics of each instrument used in this thesis will be presented in the next chapters. Analytical characteristics for each separation technique (LC, GC and CE), and each detection method mentioned (NMR, MS, UV, LIF) used in this thesis are summarized in Table 1.2, presenting the advantages and disadvantages previously mentioned, and the physical properties either measured by each detection method or used for separation.

Table 1.2. Analytical characteristics applied in metabolomics investigations.

	Physical property	Analytical technique	Advantages	Disadvantages
Separation technique	Polarity, molecular size	Liquid chromatography	▪ Fast, high–resolution separation (UHPLC) ▪ Detection of soluble thermo labile molecules ▪ Suitable for different groups of compounds (from polar to non–polar)	▪ High solvent consumption ▪ Loss of highly polar compounds
	Charge	Capillary electrophoresis	▪ High–resolution capacity ▪ Low solvent and sample consumption	▪ Most buffers non–compatible with MS ▪ Retention time shifting ▪ Only for polar compounds that can be charged
	Volatility	Gas chromatography	▪ High–resolution capacity ▪ Reproducible retention time and fragmentation pattern	▪ Thermally stable and volatile compounds ▪ High molecular weight compounds ▪ Extensive sample preparation for derivatization process

Detection method	hν absorption	UV spectroscopy	▪ Salt–insensitive	▪ Non–specific ▪ Non applicable for metabolite identification
	hν emission	Fluorescence spectroscopy	▪ High sensitivity and specificity	▪ Specific for compounds with native fluorescence ▪ Sample labelling could be needed ▪ Non applicable for metabolite identification
	Spin rotation	NMR	▪ Identification of complex structures of macromolecules (DNA, RNA, proteins) ▪ Structural information ▪ Suitable for fluids and solids. ▪ Little or no sample preparation, unbiased, rapid, robust and high reproducible, quantitative, and non–destructive to sample	▪ Low analytical resolution, overlapping signals, requires sophisticated spectral editing pulse sequences ▪ Low sensitivity (µMol and above) ▪ Large sample volume
	Mass/charge ratio	Mass spectrometry	▪ High resolution and mass accuracy ▪ Identification and quantitation of low to high molecular weight metabolites ▪ High sensitive (detection levels from ng/mL–pg/mL) ▪ Structural information	▪ No separation of isobaric, isomeric compounds ▪ Ion suppression depending on matrix ▪ Salt–sensitive ▪ Sample destructive ▪ Detect only ionized compounds

1.3.4.3. Multidimensional separations and multiplatforms

The insight research in the metabolome requires different methods and analytical platforms to capture most of the compounds in a complex mixture[11]. There are different strategies to follow when the objective is to cover more extended metabolite coverage. One approach is through the use of 2D multidimensional methods which employ two types of chromatographic columns with different separation modes which connected in parallel allow high resolution of different classes of metabolites in a single run. Multidimensional liquid separation systems (LC X LC) pose a promisingly powerful approach to enrich, separate and quantify a large variety of compounds in complex biological samples, with a powerful separation ability, high resolution and sensitivity, high peak capacity, and excellent detectability

in comparison to one–dimension LC[79]. In principle, LC X LC allows selective pre–fractionation and subsequent analysis of the target compounds[80]. However, so far data treatment is very complex and a limiting step.

In the same way, multidimensional analysis GC × GC gradually shows some advantages using columns with different selectivity enhancing the resolution power of the chromatographic system with a very sensitive detection[28, 79]. Nevertheless, these kinds of strategies still require many improvements. They still have low reproducibility making their use difficult in metabolomics. To date, there are some examples of their use in metabolomics, e.g. in the field of lipidomics analysis 2D LC–MS has been applied[58] and for 2D GC–MS different applications have been reported[81].

The second strategy, and probably the most followed by the research field, is to employ two or more analytical platforms in one study. For example the use of NMR and CE–UV (Chapter 4) which is a part of this work, shows a complementary tool to describe an animal model, combining the profiles from each technique and working with a single data matrix. In addition, as previously mentioned, profiles from GC–MS and NMR have been analysed separately and together, and the results from two independent techniques such as: LC–MS and GC–MS, CE–UV and LC–MS, etc., have contributed to the research. Very recently, a multi–platform for metabolic non–target analysis of lung tissue has been published: here LC–MS, GC–MS, CE–MS and NMR were applied and analysed independently, with the advantage that same tissue extraction was measure in all[82].

1.3.5. Data treatment

The outcome of a metabolomic study lies in the quality of the acquisition of the analytical data and in the run performance. Despite the significant advances in analytical technologies, the discovery of biomarkers in biological fluids remains a significant challenge, partly because of the overwhelming task of data treatment. The key of data treatment is the ability to distinguish genuine biological variation and metabolic changes from analytical and random interferences. It is therefore essential that data processing is sufficiently robust to enable to analyse and interpret metabolomics data properly[83].

After the metabolic "*non–targeted profiles*" are obtained they can be subjected to statistical pattern recognition techniques to reduce the dimensionality by further comparison of samples and identification of characteristic metabolic variations specific to physiological processes. Data analysis in metabolomics studies is generally extensive and there is not a standard procedure to follow. This is because all metabolomic studies yield complex multivariate data sets, which usually require visualization software, chemometric and bioinformatic methods for interpretation, and production of biomarkers that are of diagnostic or other classification value[40]. Data complexity exists because of the large number of metabolites measured simultaneously, compounded by the multiple sources of variation that can influence metabolic profiles, whether biological, environmental, or technical. For example, the nature of NMR adds to data complexity such as peak overlap and positional frequency variation[84].

1.3.5.1. Pre–processing treatment

Both data pre–treatment and processing methods depend on the analytical technique applied. ^{1}H–NMR spectroscopy and CE–UV data are typically presented on one vector–wise (i.e. one column matrix per each sample), and the data from several spectra can be organized in a 2D matrix. While the raw output of the single hyphenated MS sample for a single sample is a matrix of mass/charge ratio (*m/z*) versus retention time versus ion current or intensity, here the comparison of the samples is a 3D data set.

Correct pre–treatment of metabolomic data is an essential step before the application of multivariate statistical methods. Figure 1.5 illustrates the data treatment steps for both types of data: here the comparison of NMR & CE–UV against hyphenated–MS techniques shows examples of the instrumental outcome, the pre–processing steps and the data alignment & normalization. Each step the software used for this thesis is shown (the figure content is explained in depth in is section).

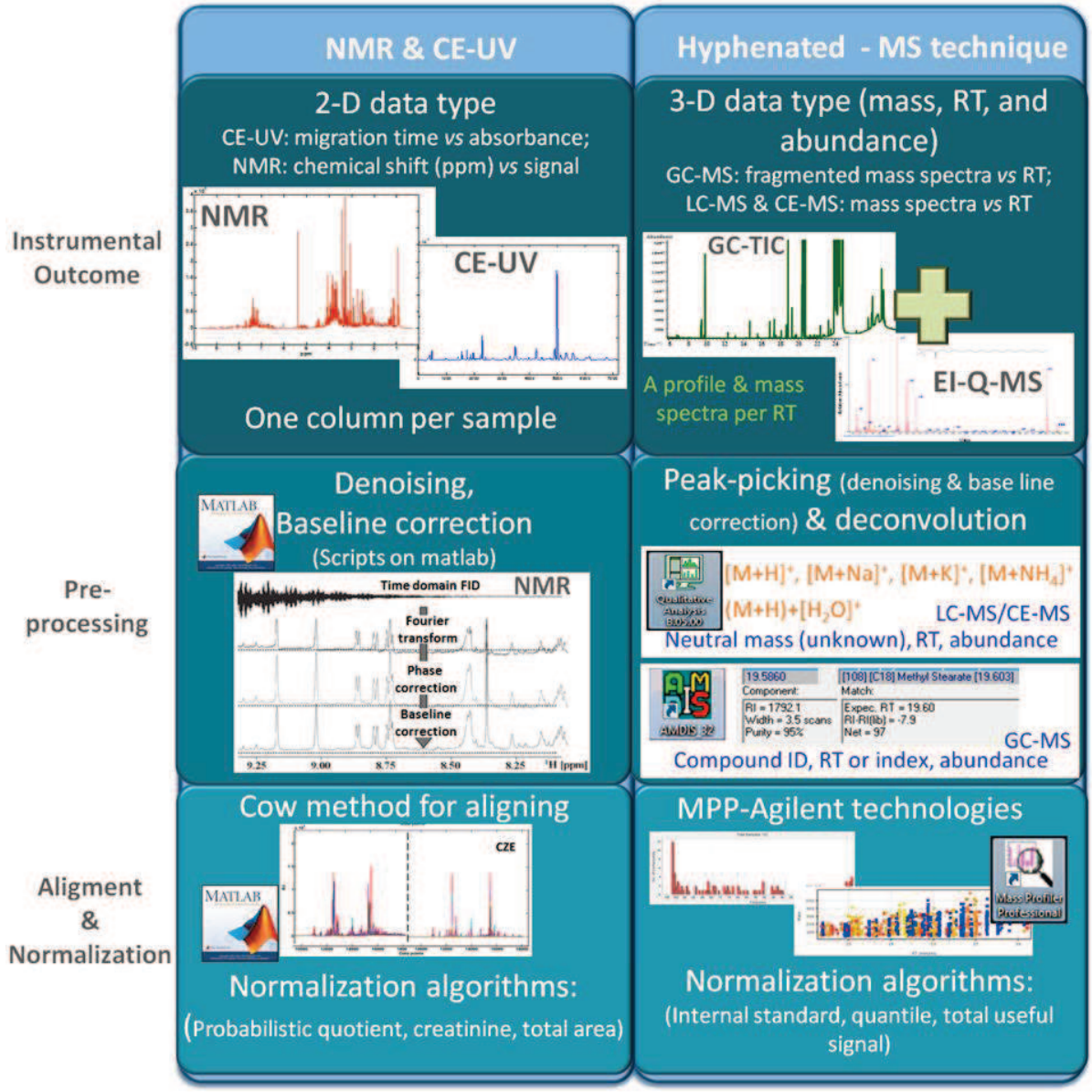


Figure 1.5. Pre–processing steps for metabolomic data. Comparison of NMR and CE–UV against hyphenated–MS techniques. General aspects about the instrumental outcome, pre–processing steps, and data alignment & normalization are simplified.

According to the two types of data that can be acquired from different analytical platforms before pattern recognition, the pre–processing data treatment consists of:

1. *2D DATA (Spectroscopic data from NMR and CE–UV).* The data treatment works managing the complete profile for each sample which generally requires several steps prior to any statistical analysis. This is because normally during the analysis of complex biological samples background noise, sensitivity decreasing over time, and retention time shifting can be enhanced. Thus, for spectroscopy–based data (NMR and CE–UV), de-noising, baseline removal, alignment and normalization, are the common followed pre–processing data steps.

Data pre–processing technologies in metabolomics based on NMR aims at reducing variances and influences interfering with further data analysis. For acquired NMR spectra, phase and baseline correction can be done manually. In addition, spectra needs to be referenced to a chemical shift, usually the methyl resonance of lactate at δ 1.33, or to the α–glucose anomeric resonance at δ5.23 is used for plasma spectra and the peak of internal standard TSP (trimethylsilyl–2,2,3,3–tetradeuteropropionic acid) at δ 0.00 for urine samples. Afterwards, spectral regions which contain non–reproducible information, or that are dominated by drug metabolites are not of interest in metabolomics, and thus need to be excluded. Typically water, and in the case of urine, urea and TSP peak resonances are removed from the data matrix. In the case of urine, the amplitude of the urea signal is falsified due to proton exchange with water.

Spectroscopic–based data normally contains both chemical and random noise. Chemical noise is caused mostly by the molecules of the buffers and solvents, while the random noise is mainly attributed to the detector[85]. In the case of CE–UV, normally the instrumental digital signal processing algorithms are efficient to properly achieve signal de-noising, such as the Savitzky–Golay implementation. In CE–UV, the baseline results mainly from temperature variations during electrophoretic runs, or the presence of impurities in the chemical composition of the buffer. Reported attempts to correct the baseline drifting in CE via mathematical algorithms includes the estimation of a new baseline by linear interpolation which was used in this work for CE–UV data. These chemometric data pre–processing methods commonly had been developed through in–house Matlab functions.

In NMR, small changes of individual chemical shifts in signals of the same metabolite in different samples is a big obstacle for data analysis and peak assignment of NMR signals in biofluids, especially in urine. Variation of chemical shifts is caused by a change of pH, salt concentration and composition and

dilution. These effects are highly reduced by buffering. On the other hand, for CE–UV, the variations in migration time – a function of electro-osmotic flow (EOF) inside the capillary, sample loading, wall interactions and physical errors (such as injection irreproducibility or temperature variations) – lead to poorly reproducible data and preclude their appropriate interpretation. In this sense, several mathematical algorithms for alignment of spectroscopic–based profiles were published to overcome this step. Nevertheless, the two aligning methods used in this thesis were correlation optimized warping (COW) method[86] for CE–UV data (in–house Matlab script developed at CEMBIO) and recursive segment–wise peak alignment (RSPA)[87] for NMR (in–house Matlab script by Velselkov, K. at Imperial College).

Finally, not all samples or biofluids are homeostatically controlled (particularly urine). Working with these differences in concentration will overlap the real changes related with the objective of the project, causing bias in the experimental result. Data normalization is usually applied in order to make the data from all samples reliably comparable with each other, and thus remove or minimize the effects of dilution in the samples[88]. It is obvious that without normalization the difference of the overall concentrations obscures the fact that only few and small differences of relative concentrations between two samples exist[89]. For example, urine volume is affected by drugs, toxins, treatments or water uptake, which increase or decrease the volume, and so metabolite concentration. It has been reported that the urine of humans and animals typically varies in concentration by a factor of 4–5, but in case of disease, food deprivation, or drug treatment (e.g., diuretics), concentration differences up to a factor of 10 between normal subjects and affected subjects have been observed[90].

Several normalization algorithms had been applied in the past, normally achieved by multiplying with a calculating scaling factor (which represents the dilution). These are: a) setting maximum peak height to the same value for all samples; b) dividing each signal value by the total area of the profile; c) dividing each signal by the value of creatinine; and the last introduced d) probabilistic quotient normalization (PQN)[89, 91]. PQN method scales spectra on the basis of the most probable dilution, which is estimated by analysing the distribution of the variables of the spectrum to be normalized, with those of a reference spectrum obtained by the median from all samples[89]. PQN is becoming the most extensively used normalization method in metabolomics. In the case of creatinine normalization, it is worth mentioning that this was the first method applied in metabolomics and was originated from clinical chemistry, where commonly urinary parameters were indexed to the concentration of urinary creatinine.

The assumption behind this was the constant production and excretion of creatinine in urine. However, it has been proved that creatinine in urine can be modulated as a metabolic response due to nephrotoxicity[92].

2. *Mass spectrometry (MS) data.* In the last few years, a large number of chemometrics tools have been developed, leading to data reduction, summarization and visualization previous to multivariate analysis[93]. In MS–based metabolomics, the starting point for data processing is a set of raw data files, each file corresponding to a single sample. Mass spectrometers operate by ion formation, separation of ions according to their mass/charge ratio (*m/z*), and detection of separated ion species. Hyphenated MS–based data will be stored as a raw data file with the instrument specific extension. These kinds of files will contain complex multidimensional MS datasets, so they need to be re–processed and converted into organized data matrices[83]. This type of data is pre–processed with a range of software to construct a 3D matrix of extracted molecular entities with related retention time, mass (or fragmentation spectrum) and abundance for each one.

For MS–based data, two types of data and mass spectra can be described: *a*) from soft ionization sources (i.e. ESI), where generally the metabolite is charged by proton $[M+H]^{+,}$ or as an adduct by other ions, i.e. with sodium $[M+Na]^{+}$ and potassium $[M+K]^{+.}$ Here little fragmentation is expected and observed in fragile molecules. *b*) MS–spectra from hard ionization sources (such as EI and CI). In this type of ion source, sufficient energy is provided to the molecules to break the bonds, producing fragment ions that have *m/z* ratios less than that of the molecular ion. Consequently, the resulting mass spectrum consists of the molecular ion peak and fragments. For LC and CE, a soft ionization source is used inside the MS, normally connected to high accuracy mass analyser such as TOF, QTOF, FT–ICR or by QqQ (this last with low mass resolution power but high sensitivity). For GC, hard ionization is coupled with either Q–mass analyser or by QTOF (for high mass accuracy of the fragments). In general typical data pre–processing pipeline goes through feature detection, alignment, normalization and filtering. These steps are discussed next.

For *soft ionization source*, a range of different software has been created to analyse MS data. Typically, to detect peaks in each chromatogram, specialized filtering and peak detection software is

used. A range of software is available from instrument companies (e.g. Agilent®, MassHunter; Waters®, MarkerLynx; ThermoFisher®, SIEVE; Applied Biosystems®, MarkerView; Shimadzu®, Profiler AM +; and LECO®, ChromaTOF) or as open–source and freely available software (e.g. XCMS[94], MZmine[95], Metalign[96] and MathDAMP[97]). For the accomplishment of this thesis the software employed for dataset creation was from the MS–instrument company Agilent Technologies®. MassHunter Qualitative Analysis Software through the algorithm called Molecular Feature Extractor (MFE) allows the creation of chemical entities (defined as "features"). The purpose of feature detection is to identify all signals caused by true ions and to avoid detection of false positives. MFE algorithm uses accuracy of the mass measurement to group ions related to charge–state envelope, isotopic pattern distribution, peaks above threshold, and presence of adducts, dimmers, and the potential loss of neutral molecules such as water to create features in the format of "*neutral mass @ retention time (RT)*" pairs. Finally, MFE algorithm finds co–eluting ions that are linked, and sums up all ion signals into one abundance value. As a result, an MFE extension file is created containing all components represented by the full mass spectral data for each sample. The final data have a 3D structure characterized by "*neutral mass, RT, and abundance*".

The obtained MFE files are then combined into a single matrix by aligning the features after re–processing by MassHunter. Features with the same mass and RT (within a mass/RT window) from each data file in the dataset are aligned along with their associated intensities[98]. For example, for LC separation, RT drifts can be the result of (*i*) variation in the temperature, pressure and mobile phase (composition and flow rates), (*ii*) changes in the chromatographic column, or (*iii*) sample matrix effects (variations in sample composition, e.g. solvent and salts)[83]. Feature alignment of MS–based data in all studies presented in the following chapters was performed using Mass Profiler Professional (MPP) from Agilent Technologies®. Special concern and care has to be taken to properly adjust the mass and RT window for feature alignment. For accurate mass analysers such as TOF and QTOF a mass window should not be higher than 20 ppm of difference.

For *hard ionization source*, normally metabolite identification is a previous step for the statistical data treatment compared to soft ions source. This is because EI ion source provides standard fragmentation of molecular ions where structural information can be performed through the interpretation of fragments ions and fragmentation patterns. This advantage is highly exploited comparing the mass spectral of a compound against spectral libraries of standards for identification prior statistical analysis. This is why

big efforts have been made to create metabolomics–specific mass spectral libraries. To date, the largest general spectra libraries available are from Wiley (~400,000 spectra) and the National Institute of Standards and Technology (NIST) (~200,000 spectra). However, mass spectrum still represents partial information, which could introduce false–positive identifications, especially for isomeric molecules. Nevertheless, hard ionization source is typically founded coupled with GC instruments, so in order to increase the identification confidence, RT information is combined with the mass spectrum matching score[99]. Thereby a specific parameter called retention index (RI) proposed by Kovats (1958) was introduced and is now used as a standardized value in reporting GC data. The Kovats' system was based on the approximate linear relation between the adjusted RT and the carbon number of *n*–alkanes[100]. The Kovats' indices are routinely used in GC as references for identification of compounds[33]. A specific library based on metabolomics and commercially available is the Fiehn GC–MS metabolomics RTL library from Agilent Technologies®, which contains approximately 700 spectra and RIs for each compound. This library was developed by Dr. Olivier Fiehn and contains available EI spectra and retention time indexes from common metabolites found in biofluids. The advantage is not only the fragmentation spectra from the standards, but the possibility to use retention time locking method (RTL), which allows for a confident identification. The RTL method is a new option developed by Agilent Technologies® which has the ability to match very closely chromatographic RTs in any GC–Agilent brand system (model Agilent 6890) to other GCs from the same brand and nominal series by calculating and adjusting the inlet gas pressure to achieve RTs of the internal standard match.

In the development of this thesis, in Chapter 2 where EI ionization was used, the compounds were identified by comparing their mass fragmentation patterns with those available in the NIST08 and Fiehn RTL mass spectral library. The typical chromatogram outcome contains hundreds of peaks of metabolites, with some compounds co–eluting, or with multiple products from reagents of the labelling process. Normally, de–convolution software uses pure mass spectra to define chromatographic peaks including overlapping peaks. Therefore, software such as Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) (which is free) was employed for automatic peak detection and de–convolution. Metabolite identification could be achieved by several strategies, however RI scoring and RT analysis are the most often applied. The ID provided is obtained by matching the RT or the RI and the mass spectrum of the sample peak with those of a pure standard previously analysed under identical

instrumental conditions (in our case, the Fiehn RTL method)[27]. AMDIS give as a result a file with all identified compounds (with related RT or RI, EI fragmentation mass spectrum and peak area).

As an example of these two type of ion sources, Figure 1.6 shows the spectra of stearic acid in a soft ion source (Figure 1.6A) by ESI in LC–QTOF–MS in positive mode, and hard ion source (Figure 1.6B) done by EI in GC–Q–MS. In the ionization by ESI, adduct formation with sodium (Na^+), potassium (K^+), ammonium (NH_4^+), and the loss of water molecule can be observed. The outcome of the compound identification is based on the accuracy of the mass analyser, which for this example was a QTOF. For hard ionization, typical pattern fragmentation is observed where the base peak is not normally the highest mass peak, and of note this is a–trimethylsilyl derivative. Also in the spectrum typical fragments such as 73 and 147 are well known to show up from reagent (trimethylsilyl groups).

The work in AMDIS is more extended to achieve data–treatment. In detail, a standard mix formed with different fatty acid methyl esters (FAME) is analysed along the non–targeted analysis work–list. Normally, 10 compounds are chosen covering most of the RTs from the chromatogram. Then, FAME entries (with the experimental RTs of FAME updated) are used to create a calibration data file for further RTs adjustment in the samples. This improves accuracy in the determination of the experimental Kovats retention index (RI). Peak detection and deconvolution is performed by the algorithm: "*Use Retention Index Data*" in AMDIS[101]. Then the well–established RI value contained in the Fiehn RTL library is compared to the experimental RI value and fragmentation pattern after deconvolution to assign a net match score between the experimental RI and the theoretical one[102]. Deconvolution is optimized by a second analysis performed by the "*Use of retention time*" algorithm in AMDIS. Those metabolites not included in the Fiehn RTL library (e.g. cholesterol and arachidonic acid) can be added from NIST library. File data generated by AMDIS can be aligned by MPP software; the alignment is not only performed by the compound ID but by the number of qualifier ions matched with the spectrum of the standard.

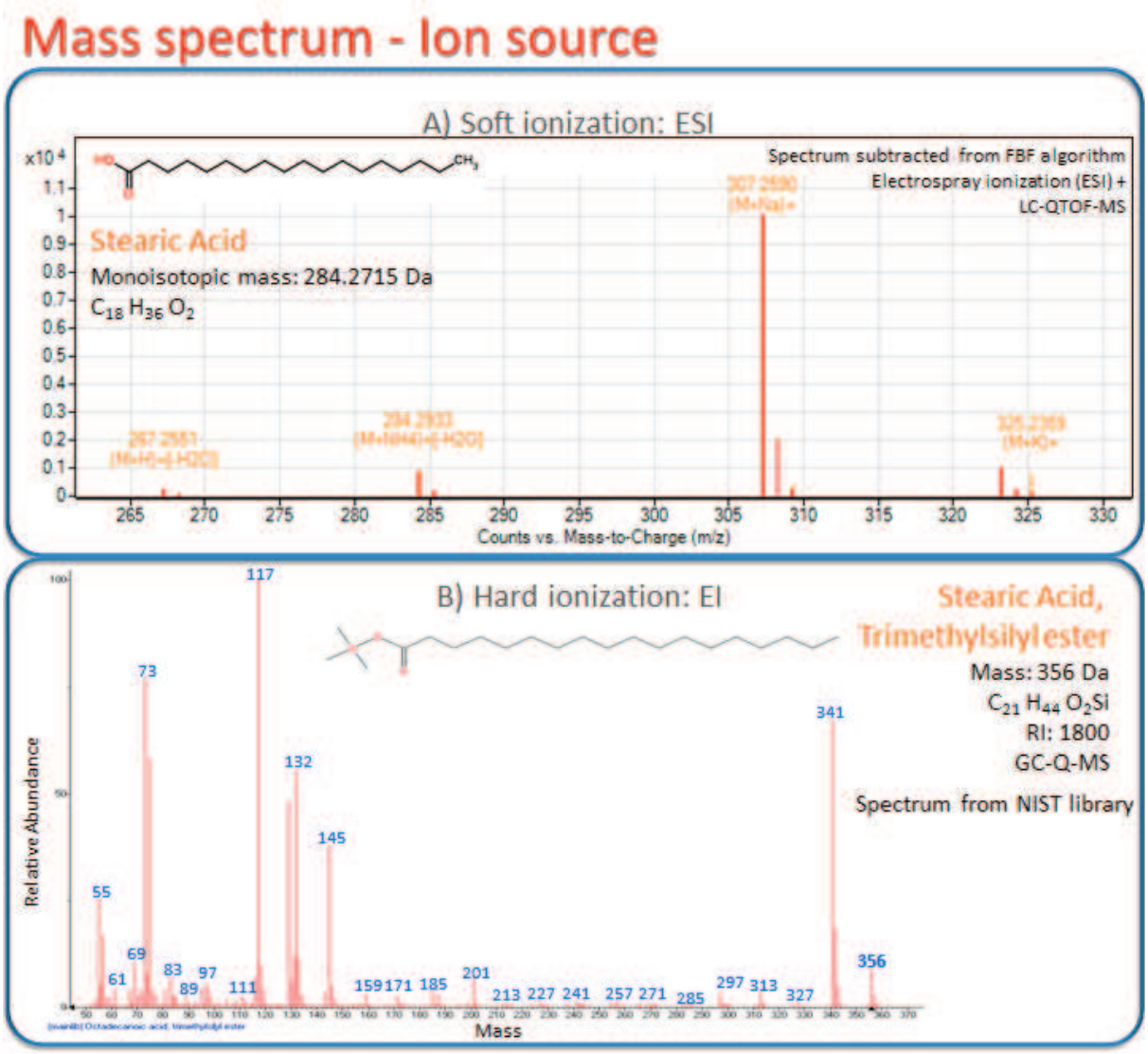


Figure 1.6. Mass spectrum example according to the type of ion source for stearic acid. A) Soft ion source: electrospray (ESI) in positive mode from LC–QTOF–MS (subtract spectrum from Find by Formula (FbF) algorithm); B) Hard ion source: Electron impact (EI) from GC–Q–MS for stearic acid (analytical conditions of both instruments are in the article of Chapter 2).

In comparing NMR or CE–UV, data normalization in MS–based techniques do not have a complete establishment of an algorithm. This is because of the presence of missing values in the data matrix and the drift observed in the abundance. Several normalization strategies for MS–data have been

proposed. Among them, MS–total signal (MSTS), MS total useful signal (MSTUS), locally weighted scatter plot smoothing (LOESS), and quantile and median fold change, all reviewed in[93, 103]. Here additional steps for data pre–treatment can be added, such as treatment of missing values or data logarithmic transformation to solve data drift.

The challenges faced with the data generated by MS are several: first, the analysis is complex because the number of variables (features) is much higher than other techniques (especially for LC–MS) with a relatively small number of observations, large contribution of noises and biases beside biologically relevant information, many missing values or drift in the analytical response. However data filtering and normalization (in case to be needed, i.e. for urine) are part of the pre–treatment steps to overcome these situations. Data filtering in MPP can be performed through different strategies. In the studies performed in this thesis, two filtering steps were applied in this order: first, a "*filter by presence*" was performed prior to the "*filter by sample variability*". Filtration according to sample variability is carried out based on coefficient of variation (CV %). Filtration is normally performed by keeping the compounds present (from 80-100%) in at least one group, and then from obtained matrix masses that have a CV above 35% on quality samples are filtered out.

When the data matrix is complete and the pre–treatment processes have been applied, the next step is the statistical analysis of the data.

1.3.5.2. Multivariate and univariate analysis

The advance in analytical techniques in the metabolomics field has resulted in the generation of large, complex datasets, which in turn has led to a demand for improved data analysis approaches, not only for preprocessing steps but for advanced chemometric approaches too[104]. The large amount of information obtained from the analytical techniques does not ensure that statistical differences will be present. Moreover, the incorrect statistical treatment may result in producing misunderstandings and spurious results, and in overlooking important information. Once the data have been collected and it has been preprocessed in form of a data matrix, statistical analysis is performed. Two types of approach can be described: univariate and multivariate analysis.

1. Univariate approach.

In this setting, the association of each variable and the outcome of interest (e.g. case/control studies) is assessed and tested using the same statistical model, leading to as many tests as there are variables for each outcome[105]. The typical statistical models are based on linear regression, parametric tests (two–sample groups *t*–test or analysis of variance (ANOVA) for more than two groups) and non–parametric test (Mann–Whitney *U* test for two–sample groups and Kruskal–Wallis for the non–parametric ANOVA data). A decision tree is observed on Figure 1.7. For *t*–test, data must follow normal distribution while for ANOVA, normality and homogeneity of variances are compulsory requirements.

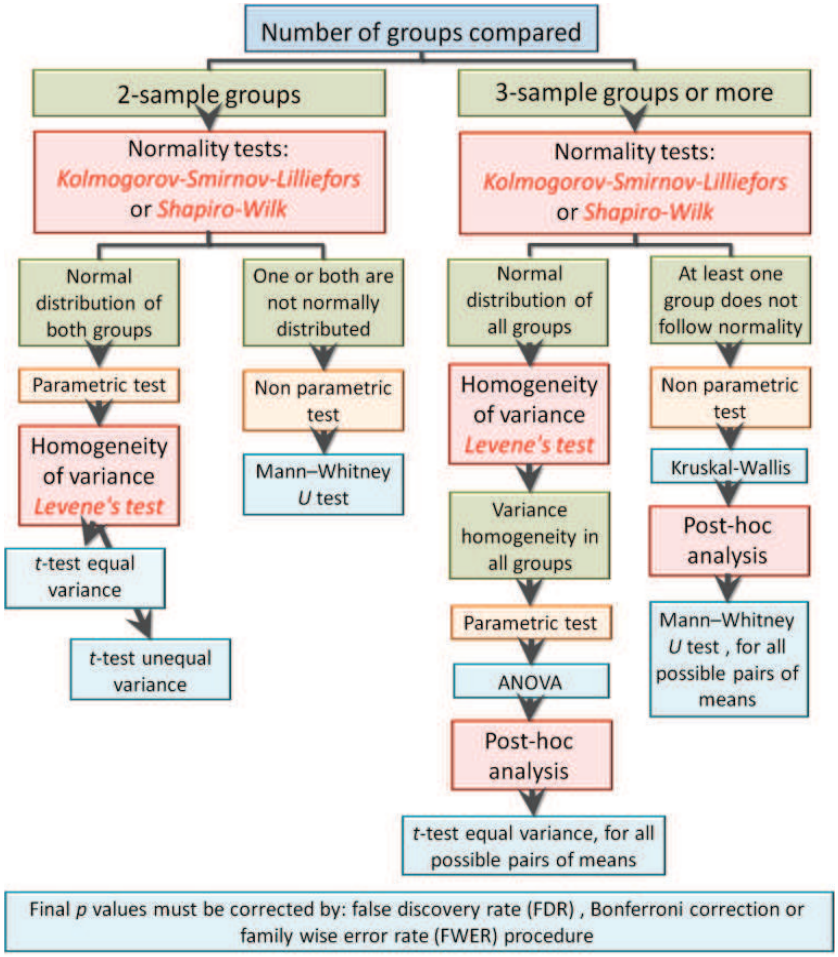


Figure 1.7. Decision tree for univariate statistical models. For two–sample groups: *t*–test and Mann–Whitney U test, while for more than two groups: ANOVA and Kruskal–Wallis can be applied. (see conditions) in the diagram.

For all these univariate models the strength and significance confidence is often 0.05. Thus, when several hundred or more tests are performed, the overall number of false positives is increased, leading to a greater risk of error. In order to avoid the inflated number of associations that are falsely declared as statistically significant across all performed tests, several corrections have been proposed: False Discovery Rate (FDR); Bonferroni correction; and the family wise error rate (FWER) procedure. Of all, the most popular method is the FDR. This was developed by Benjamini and Hochberg in 1995, and in this correction only 5% of the statistically significant variables could be false positives on average[105].

Univariate methods are useful to uncover simple relationships between variables and responses. This approach offers the advantages of: (*i*) simplicity and flexibility, as they can accommodate a large range of parametric and non–parametric relationships; (*ii*) they are computationally efficient, as they can be applied to large datasets (i.e. through a Matlab script); (*iii*) these methods are readily available in most statistical packages. Unfortunately, since they consider each variable independently, they are unable to identify general patterns of correlation among the variables, which can lead to redundancy and unnecessary complexity in the results. More importantly, univariate methods are also likely to ignore relationships involving combinations of variables[105]. Therefore univariate analysis contributes to the statistical process but multivariate methods are highly important for data treatment in the metabolomics field.

2. *Multivariate methods.*

Metabolomics data are multivariate, so multivariate techniques should be used to analyse the multiple variables simultaneously. This is because the output table (even after data filtration) is a large matrix of dimensions: *N* (samples) * *K* (variables), here normally *N* is small while *K* is typically tens to hundreds or thousands of metabolite features or variables[104].

For more than a century (Pearson, K., in 1901), multivariate analyses have been used to extract possible complex structures from high–dimensional data[105]. Therefore different multi–dimensional and multivariate statistical analyses and pattern–recognition programs have been developed to handle the large amounts of data in an effort to interpret the complex metabolic pathway information from the

measurements[48]. Chemometric analysis can be performed using various software packages such as SIMCA–P (Umetrics, Umeå, Sweden) or by using customized scripts in programming languages such as Matlab or R. Multivariate models are able to handle many, incomplete and correlated predictor variables in a simple and straightforward way. The information resides in the correlation structure of the data. These multivariate analyses are usually divided into unsupervised and supervised methods.

(i) *Unsupervised methods.*

Unsupervised methods are used to analyse the data without any *a priori* information concerning the sample groups[104]. Among the various methods, principal components analysis (PCA) has a predominant role, although there are other methods such as: hierarchical clustering, independent component analysis (ICA), multi–dimensional scaling (MDS), *k*–means clustering and self–organizing maps (SOMs). All these are outside the scope of this thesis. PCA is the workhorse in chemometrics and is the first step in data analysis. This allows the researcher to observe patterns in the data, providing a summary or overview of all samples in the study, and to identify molecules responsible for the patterns. Groupings, trends and outliers can also be found[106].

PCA is a projection–based method, where the main assumption is that the system dataset can be driven by a small number of principal components (PCs). PCs are intended to reduce the dimensionality of the dataset to a more manageable size, thus improving its accessibility, and streamlining data interpretation. However, these PCs might not offer a particularly good explanation of the response variables[105–107]. Linear combinations of all peak intensities are constructed to produce orthogonal components that maximize the total variance in the samples, independent of their group labels[48]. One way to explain PCA is depicted in Figure 1.8 (adapted from Trygg, J., 2007[106]), from the typical dataset composed by *N* samples and *K* variables (represented here for only 3 variables) they can be projected in a plane of two dimensions. The first principal component (t_1) in PCA model describes the largest variation in the samples as possible. In addition, from PCA, a score plot and loadings plots are obtained. The position of each sample in the model plane is used to relate variables to each other (scores plot). Importantly, the loadings measure the contribution of each original variable to each PCs explaining the position of observations in the scores plot, and thus they can be used to identify which characters differ most between groups.

PCA is typically useful to evaluate the quality of the data collected from the analytical platforms, more specifically for MS–techniques where the stability of the analysis may suffer deep drifts. For example, chromatographic–MS techniques are susceptible to degeneration over time due to several factors, such as column contamination and MS–ion source contamination. These factors can be controlled using quality control (QC) samples, i.e., the repeated injection of a randomized sample throughout analysis. QC's position in the PCA model shows the quality and the stability of the instrument performance during the experiment. Therefore, the PCA model is not suitable for selecting reliable biomarkers, because the two first components do not necessarily contain the most relevant variations between the groups. Although frequently, biomarkers identified from the loadings plot of PCA model, are confirmed by supervised models[108].

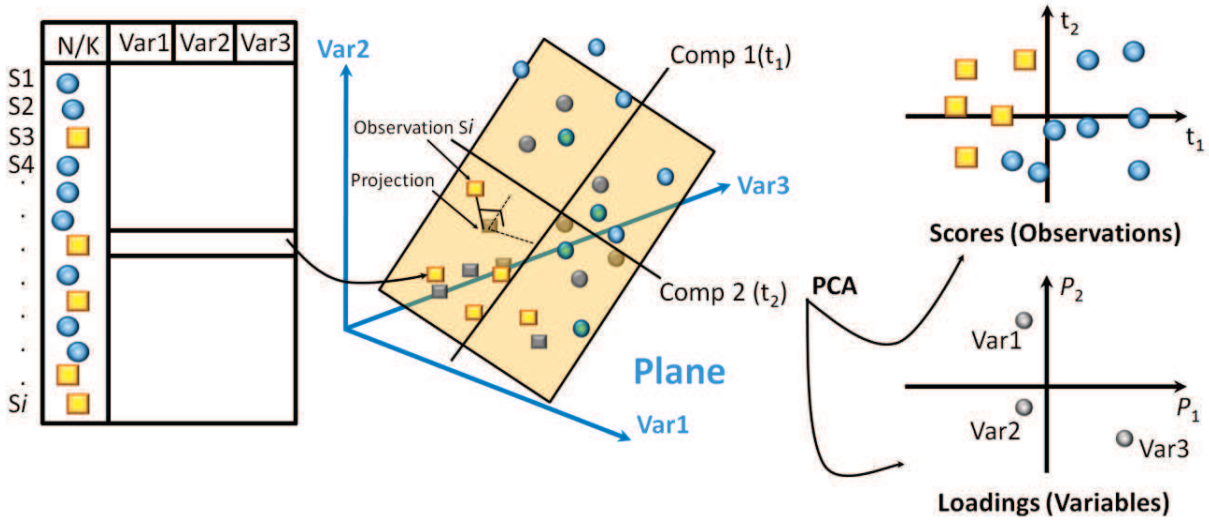


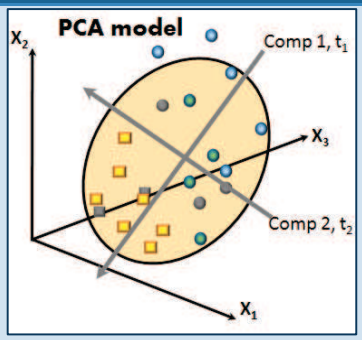
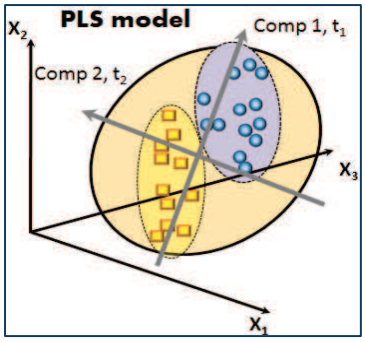
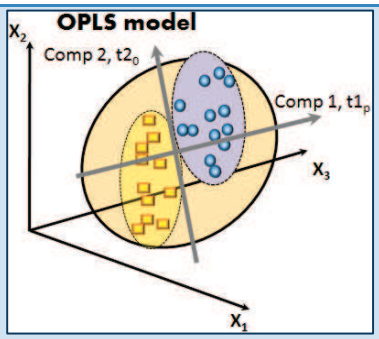
Figure 1.8. A principal component analysis (PCA) model explains the variation in the data set ($N_{samples}$ / $K_{variables}$) by low–dimensional model. The PCA model provides a score plot, where the relation among the samples in the table is visualized. Loading plot describes the influence of each variable (Var) in the model plane and the relationship among them. Importantly, directions in the score plot match to directions in the loading plot (taken from Trygg, J., 2007[106]).

(ii) *Supervised methods.*

Typical supervised techniques applied to metabolomics data include the projection to latent structures (PLS) based approach, which includes: partial least squares (PLS) regression; orthogonal PLS (OPLS); PLS discriminant analysis (PLS–DA); and OPLS–DA[109]. Other less common supervised methods are random forests, support vector machine (SVM) and *k*–nearest neighbour algorithm (*k*NN)[104]. The PLS methods can extract Y–correlated information from the data matrix (X) seeking maximum covariance model of relationship between X– and Y–variables. In the PLS-DA method the X matrix is an organized dataset, while the Y matrix is the supervised data; Y corresponds to the class which each sample belongs to. In other words, for PLS-DA approaches, the algorithm takes into account sample class information and tries to construct a model that could predict the membership of the sample. Thus, they are convenient methods for extracting meaningful metabolites correlated with the class.

A modification of the PLS method, named OPLS (orthogonal PLS), can be used to facilitate the interpretation. OPLS separates the variation in the data matrix X into two parts: one part correlated to Y and one part orthogonal and irrelevant (not related) to Y. In an OPLS–DA of two classes, for instance, all the information relevant for the explanation of the class is summarized and used as a predictor in one component, which corresponds to the class[110]. The OPLS–DA makes the interpretation of the model more straightforward by separating predictive from non–predictive (orthogonal) variation. Thus, PLS–DA models are mainly used for sample classification while OPLS–DA is used for the screening of potential biomarkers[103, 103–111]. All of the characteristics of multivariate models are summarized in Table 1.3. Here, a scores plot was depicted using a geometrical illustration for each model (PCA, PLS and OPLS). Among the characteristics described in the table are some of the aspects such as validation and biomarker identification, which are explained next in this chapter.

Table 1.3. General characteristics of multivariate models PCA, PLS–DA and OPLS–DA, along with a geometrical example of scores plot.

Multivariate model	Characteristics	Scores plot example
Principal Component Analysis (PCA)	▪ Trends & patterns ▪ Strong outliers by hotellings T2 plot & moderate outliers by DModX plot ▪ Quality control checking ▪ Patient monitoring (time trajectory) ▪ Biological diversity ▪ No real predictive ability	
Partial Least Square Discriminant Analysis (PLS–DA)	▪ Discrimination between groups is explained by the components (e.g. t_1& t_2) ▪ Sample classification (two or more groups) ▪ Predictions ▪ Biomarker candidates (variable importance (VIP), loading plot) ▪ Validation is needed (CV1, CV2, permutation test)	
Orthogonal PLS–DA (OPLS–DA)	▪ Discrimination between groups is explained by first component (class= t_1) ▪ Classification forced ▪ Predictions ▪ Biomarker candidates (*S*–plot, *SUS*–plot, jack–knife) ▪ Validation is needed (CV1, CV2)	

Comp 1, $t1_p$ is the predictive component while Comp 2, $t2_0$ is the Y–orthogonal component. Scores plots were adapted from Trygg, J., 2007[106].

Before performing any multivariate model, data is usually scaled. This is essential to remove the influence of the variable averages and focus on the variation among the data. If scaling is not performed, the analysis will focus only on metabolites with high intensities, as they are likely to show higher variance than variables with low intensities[104]. Among the scaling more often used in metabolomics, unit variance (UV), mean centre (Ctr) and Pareto are the most common. For UV, each variable is divided by its standard deviation, resulting in a variance of 1 for all metabolite features. This gives the same weight to all variables in the dataset, and so can be risky if the dataset contains lot of noise, as this will be as interpreted as real data. For Pareto, scaling the square root of the standard deviation is used. This reduces the relative importance of metabolites with high intensities but gives less weight to the small intensities. Finally, for Ctr the average value of each variable is subtracted from the variable. Therefore, Ctr adjusts the differences between high and low abundant compounds.

Moreover, for MS–based data, logarithmic transformation has been widely applied before modelling. This is due to the high variability and dispersion of values inherent to the MS–data technique. The aim of logarithmic transformation is to reduce the influence of potential outliers, and also to transform the data matrix into a more normal–distribution type. This is used to correct the heteroscedasticity and is also considered pseudo scaling[112]. Moreover, logarithmic transformation increases the weight of low–intensity metabolites and compresses the upper end of the measurement scale[93]. A drawback of the log transformation is that it is unable to deal with the value zero.

All of the multivariate models need to be validated and, to have an understanding of their validity and limitations. It is possible that the good statistical performance observed in the chemometric models can be due to over fitting of data[113]. There are several strategies to achieve this purpose. First, whenever a model is constructed, the software calculates a couple of quality descriptors (Q^2 and R^2). R^2 explains variance, and gives a score about the goodness of sample classification, while Q^2 predicts variance, providing information about how well the model can predict a new sample. These two scores (R^2 and Q^2) should be high and not vary more than 0.2–0.3 between them, and they are considered as the internal cross validation of the model (CV1). Other typical ways to validate PLS models is through the permutation test, which compares the original quality scores of the model (R^2 and Q^2) with those obtained after randomly permuting the samples to a different class, expecting to get worse quality scores (even negative for Q^2), while more samples are permuted to a different class[110, 114]. Once the validity of the

model is confirmed by internal validation strategies, data can be further confirmed using external validation (CV2). Here the samples are divided into two groups, one called training set (i.e. 2/3 of total samples) which is used to build the model and prediction set which will be the group of samples left used to be predicted by the model[113].

1.3.6. Biomarker peaking and identification

After multivariate models are performed (PCA, PLS, OPLS) it is important to evaluate the quality of each one by the tools previously described. It is necessary before selecting the compounds that are responsible for group separation that the models are of good enough quality to obtain reliable results. Normally in the comparison of two groups (e.g. healthy vs disease), samples ideally should have good classification pattern in PLS–DA, and good quality scores for the OPLS–DA model in order to look up significant variables. Biomarker peaking is part of the multivariate workflow that can be done through the use of several approaches (Figure 1.9):

1. *Loading plot with jack–knife confidence interval*. The most common is through the use of the OPLS–DA model. This is because, as previously mentioned, OPLS–DA had the advantage that a single component is used as a predictor for the class improving interpretation. Here the use of a loadings plot in format of a column (CLP) depicts the weight and correlation of each variable with the class. Each variable in CLP is always associated to a jack–knife interval, which is the error bar of confidence to be significant for the correlation. The most interesting variables are those with high correlation and covariance with the class (large height of the column and jack–knife significant: Figure 1.9A)[115].
2. *S–plot*. In the case that Pareto or Crt scaling are used for OPLS–DA modelling, the *S*–plot can be performed. This plot is another way to visualize OPLS–DA loadings combining with the modelled covariance (*p*) and correlation (*p(corr)*) in a scatter plot. In the correlation axis the "*reliability*" of each variable is represented, while the covariance axis describes the magnitude "*weight*" of each variable in the model. Ideally, biomarker peaks have high "*reliability and weight*" while peaks with low "*reliability and weight*" are close to the noise (non–relevant) level (Figure 1.9B)[115].

3. *Shared and unique structure (SUS)*–plot. This plot shows the correlation of variables when they are involved in two comparisons, each one coming from a separated OPLS model. This is specially worthy to detect the variables highly correlated with a condition when the data is under two classes (e.g. gender and treatment: Figure 1.9C) [115].
4. *Variable importance (VIP) plot.* The VIP values reflect the importance of the variable in the PLS model. The VIP values mainly reflect the correlation of the terms to all the responses and the plot displays the VIP values as a column plot sorted in descending order with confidence intervals derived from jack-knife. VIP values greater than 1.0 are predefined as important variables in the discrimination of test (Figure 1.9D)[113].

When the significant variables are selected, additional data treatments can be obtained, such as: evaluation by univariate analysis; percentage of change; and coefficient of variation in QC samples as good describers. Now, the most challenging part of a metabolomic study is confirmation of biomarker identity. This is an essential step towards understanding the biological changes occurring within the system and remains a major bottleneck in metabolomics investigations.

Identification of metabolites will depend on the analytical technique employed. In the case of NMR, experimental and mathematical tools have been used. Two dimensional experiments (TOCSY, COSY, HSQC and HMBC) offer a good option in the identification and structural elucidation of unknown compounds[48]. While statistical spectroscopy has developed several strategies, the most widely used for identification is STOCSY[116]. For CE–UV, metabolite identification can only be made via trial and error, using pure standards of compounds expected in the profile. Correlation algorithms (linear autocorrelation matrix) have also helped when the data matrix of CE–UV is combined with data from a technique with structural elucidation like NMR (CE–NMR–SHY)[117].

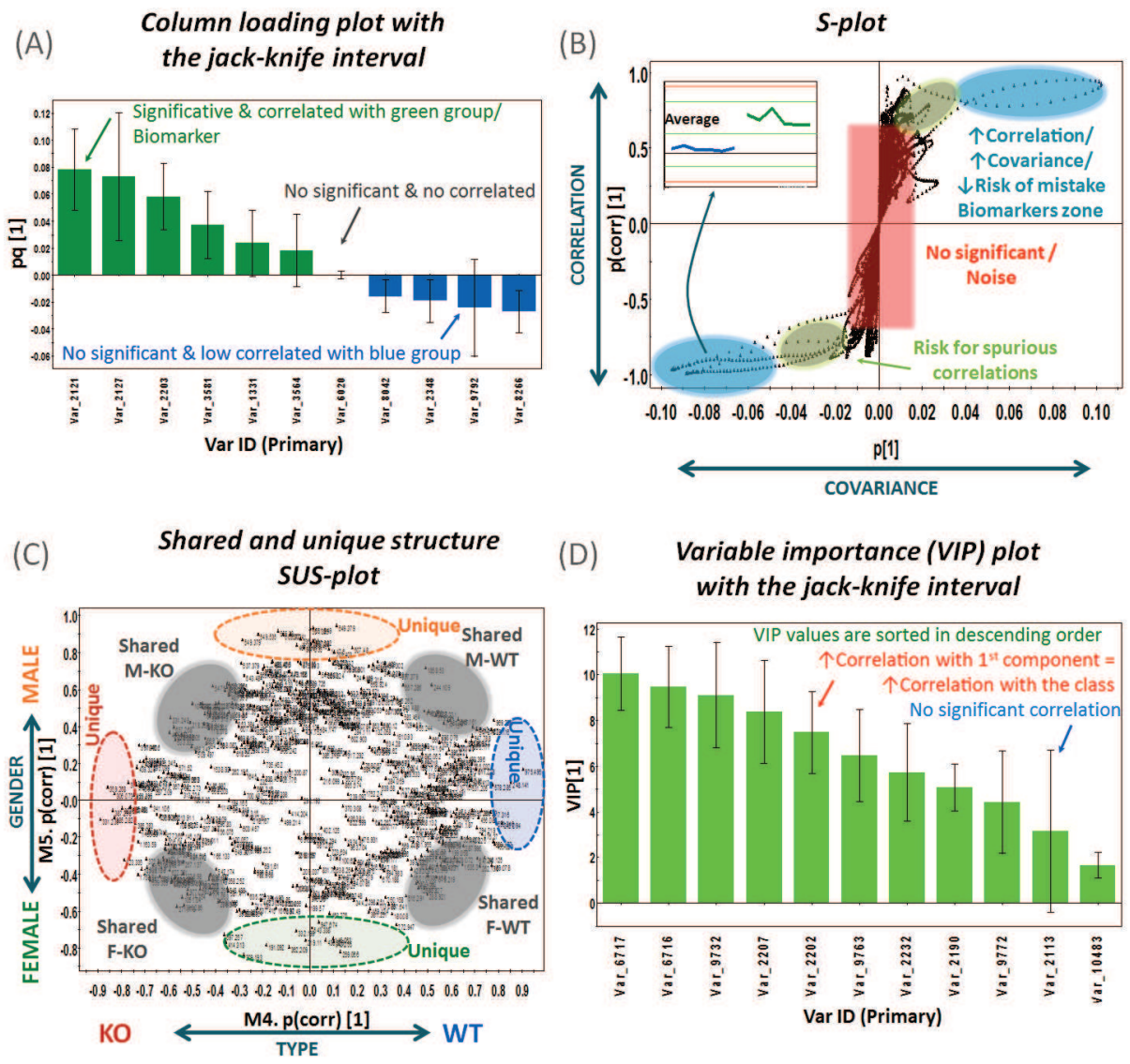


Figure 1.9. Biomarker peaking methods (examples): (A) Loading column plot with jack–knife interval confidence from OPLS–DA model; (B) *S*–plot from OPLS–DA model; (C) *SUS*–plot correlation from two independent OPLS–DA models; (D) VIP–plot of the first component in PLS–DA model (all graphics were obtained from real data using SIMCA).

In the case of MS–based data, the identification starts with a "*mass@RT*" of a compound. To tentatively identify the compounds, neutral accurate masses are searched for in several databases (METLIN, Lipid Maps, MASSTRIX and HMDB) for hits. For the later works in this thesis, the "*CEU mass mediator*" database was used only, as this was an interface mediator which automatically searches in KEGG, METLIN and Lipid Maps databases (http://ceumass.eps.uspceu.es/mediator/). For each hit, the chemical formula was compared with the experimental isotopic pattern distribution.

To really confirm the identity of metabolites, MS/MS experiments needed to be performed. These can only be carried out by MS instruments that are able to select a mass and then fragment it, such as QqQ, Q–TOF, ion–trap, and ICR–FT. Normally, MS/MS experiments are required to be performed on identical conditions and instruments as the primary analysis was performed using the same analytical conditions. The conditions for the MS have to be adjusted, for example: in the case of Q–TOF, the use of collision induced dissociation (CID) for MS/MS is needed. The quadrupole selects the specific mass and this enters the CID, where a collision gas (usually nitrogen) and collision energy is employed to break up the compound. The energy applied can be set to an specific value (10, 20 or 40 V) or could be adjusted using the interpolation of a slope of 3.6V/100Da. The ions are selected or targeted based on their *m/z* ratio in a RT window (previously determined).

Figure 1.10 shows an example. In brief, the corresponding molecular ion (*m/z*) is searched as an extracted ion chromatogram (EIC) and matched to its RT (Figure 1.10B). Elemental composition of the feature (chemical formula) based on the exact mass and isotope pattern recognition is compared with the database hit, considering the probability score. For compounds with a database hit and probable chemical formula match, structural information is obtained from the MS/MS experiments. The MS/MS spectra can be analysed comparing the compound spectra with the MS/MS spectra library if available (Figure 1.10C). Fragmentation patterns of compounds can also be predicted either by instrumental software such as: Molecular Structure Correlator from Agilent Technologies®, or by commercial software such as ACD/ChemSketch (ACD/Labs, Toronto, Canada). The advantage of using Agilent–brand software is that it uses the raw file from the instrument and bases the identification on the experimental fragments, giving a score and a weight of structural elucidation. Finally when possible, MS/MS of standards are needed to show that the standard and the unknown have approximately the same RT, isotopic distribution, and fragmentation pattern (Figure 1.10D).

However, the identification in MS is not easy. Some of the troubles found are: typically, biomarkers are in small amounts, so fragmentation and isotope distribution are not clear; some compounds present lot of fragments difficult to explain; conversely, some have such a strong structure, they had no fragmentation; and co–elution of other compounds with close *m/z* make fragmentation spectra spurious and hard to analyse. From the study in Chapter 5[93], from an initial and significant data set of 165 compounds only 20 identifications of compounds were confirmed.

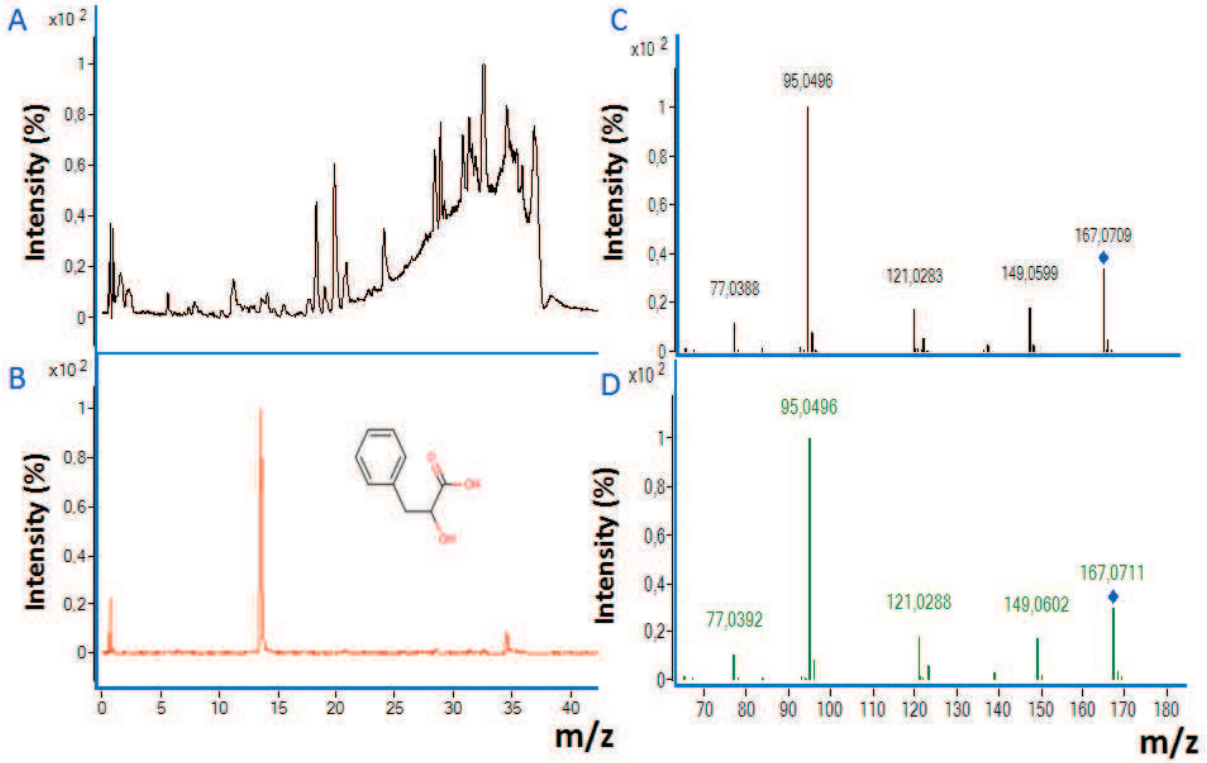


Figure 1.10. Identification of a selected biomarker using the example of phenyllactic acid (PLA)[118]. (A) Total ion chromatogram from a sample was selected. (B) Extracted ion chromatogram of 167.0696 (m/z), RT of 13.6min. (C) MS/MS experimental spectrum from the ion [167.0696 (*m/z*), RT = 13.6min]. (D) MS/MS spectrum of commercial analytical standard PLA reagent (collision energy = 175 V).

1.3.7. Biochemical interpretation

Once the compounds are confirmed their resulting changes should be explained, in the context of the case of study. If possible, these changes should be mapped to known metabolic pathways to build biochemical hypotheses. Pathway information can aid the interpretation of biological changes indicating the significant differences in metabolite concentration identified through the metabolomic approach. The translation from analytical changes of metabolites into biological knowledge represents a significant challenge for biologists and bioinformatics. This is a hard task for several reasons: for typical mammals it is necessary to consider they are not single organisms, but systems comprised of a combination of mammalian, bacterial and potential parasitic organisms. Specifically, there is symbiosis between a host organism and its gut microbiota, and that gut microbes have functions and enzymes that are not found in mammalian organisms[119]. In addition, the pathways are not separate and static; they are interconnected, and plenty of these interconnections and regulations are still unknown. Also, very frequently, the metabolites found significant are from very different pathways, giving only one single metabolite related to each pathway, which makes it difficult to interpret or to connect to other metabolites. Lastly, new knowledge about enzymes and molecular mechanisms is discovered every day. Nevertheless, there are some well–known pathway databases such as KEGG and BIOCYC, which are free to consult and, in the case of BIOCYC, it is possible to draw new pathways. For KEGG, this is an online resource where the interaction information of genes, proteins and metabolites is integrated and it can be used to investigate molecular networks in specific organisms or for all[120]. Moreover, MS–instrument software (MPP) from Agilent Technologies® has a tool for pathway analysis where it allows the visualization of created pathways from WikiPathways and BIOCYC databases, with the possibility of integrating information from articles.

1.4. Target approach

The targeted approach normally has different nuances. After the molecular mechanism and biochemical hypothesis is established, the next step should be the validation of these biomarkers. Here the researcher should be able to confirm their significance and reliability. Thus, development of a quantification method is important.

In addition, targeted analysis could be necessary for comparison with previous observations. For example in the case of Chapter 6, for bipolar disorder (BD), the imbalance of amino acids in patients with BD in comparison with healthy controls had been previously reported.. The optimization and validation of a new method specific to the quantification of amino acids in plasma is the first objective before the detection of differences between experimental groups.

Analytical methods for targeted analysis should be validated. Koek, M., in 2010[121] suggested the validation for at least the following parameters: selectivity, calibration model (linearity and range), accuracy, precision (repeatability and intermediate precision) and limit of quantification. Here, not only the analysis of amino acids in plasma in Chapter 6 but also the breast milk non–target analysis with LC–QTOF–MS and GC–Q–MS in Chapter 2, shows the complete validation format.

1.5. Concluding remarks and future trends

The research in metabolomics has evolved to a new level. The advances in instrumental equipment, informatics and software for data acquisition and treatment, and chemometric and multivariate experience have made metabolomics a very strong and reliable tool in the scientific arena. Nevertheless, there are some aspects that at the present are not totally established or studied in depth, e.g. the missing values in hyphenated MS–techniques are still some of the gaps that need to be explored. Furthermore, for any MS–based technique, the identification based on databases and MS/MS confirmation is limited. New strategies and tools are needed to be capable of identifying and characterizing new structures or compounds that are not present in a database, or whose pattern fragmentation cannot be explained.

As the workflow in metabolomics is not quick, and each step takes time and effort, it frequently happens that after the significant compounds were found and explained in a biochemical hypothesis, that they are not validated in a second set of samples. This is reasonable, as the validation of these compounds implies the development of a new method with a targeted approach for their quantification in a new set of samples, which most of the time is hard or not possible to collect.

In the same way, for animal models and cell culture studies, these results should be moved forward and expanded or tested on human models. Moreover, when the purpose of the metabolomics study is related to the population and most of the times it is, i.e. in the diagnostic field, epidemiological studies should be taken up. Population size samples should be analysed, but that should be done in a targeted way.

Along with the validation of the results in populations their application in clinics should be performed. One of the major objectives in metabolomics is its contribution to the clinic field and human well–being. So, its contribution should be followed with the creation of new diagnostic and prognosis tests. Those would be based on the measurement of a set of compounds instead of a single and classical biochemical metabolite, marking the probability, or diagnose for a specific disease.

REFERENCES

1. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* **2002,***1* (2), 153–61.

2. Lindon, J. C.; Holmes, E.; Bollard, M. E.; Stanley, E. G.; Nicholson, J. K., Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **2004,***9* (1), 1–31.

3. Fiehn, O., Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* **2002,***48* (1–2), 155–71.

4. Hocquette, J. F., Where are we in genomics? *J Physiol Pharmacol* **2005,***56 Suppl 3*, 37–70.

5. Hollywood, K.; Brison, D. R.; Goodacre, R., Metabolomics: current technologies and future trends. *Proteomics* **2006,***6* (17), 4716–23.

6. Verpoorte, R.; Choi, Y. H.; Mustafa, N. R.; Kim, H. K., Metabolomics: back to basics. *Phytochemistry reviews* **2008,***7*, 525–537.

7. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F., Systematic functional analysis of the yeast genome. *Trends Biotechnol* **1998,***16* (9), 373–8.

8. Nicholson, J. K.; Lindon, J. C.; Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999,***29* (11), 1181–9.

9. Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L., Metabolite profiling for plant functional genomics. *Nat Biotechnol* **2000,***18* (11), 1157–61.

10. Fiehn, O., Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* **2001,***2* (3), 155–68.

11. Johnson, C. H.; Gonzalez, F. J., Challenges and opportunities of metabolomics. *J Cell Physiol* **2012,***227* (8), 2975–81.

12. Westerhoff, H. V.; Palsson, B. O., The evolution of molecular biology into systems biology. *Nat Biotechnol* **2004,***22* (10), 1249–52.

13. Oresic, M., Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. *Nutr Metab Cardiovasc Dis* **2009,***19* (11), 816–24.

14. Nicholson, J. K.; Wilson, I. D., Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* **2003,***2* (8), 668–76.

15. Nicholson, J.; Holmes, E.; Lindon, J.; Wilson, I., The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* **2004,***22* (10), 1268–74.

16. Nemutlu, E.; Zhang, S.; Juranic, N. O.; Terzic, A.; Macura, S.; Dzeja, P., 18O–assisted dynamic metabolomics for individualized diagnostics and treatment of human diseases. *Croat Med J* **2012,***53* (6), 529–34.

17. Nicholson, J. K., Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* **2006,***2*, 52.

18. Mirnezami, R.; Nicholson, J.; Darzi, A., Preparing for precision medicine. *N Engl J Med* **2012,***366* (6), 489–91.

19. Roychowdhury, S.; Iyer, M. K.; Robinson, D. R.; Lonigro, R. J.; Wu, Y. M.; Cao, X.; Kalyana–Sundaram, S.; Sam, L.; Balbin, O. A.; Quist, M. J.; Barrette, T.; Everett, J.; Siddiqui, J.; Kunju, L. P.; Navone, N.; Araujo, J. C.; Troncoso, P.; Logothetis, C. J.; Innis, J. W.; Smith, D. C.; Lao, C. D.; Kim, S. Y.; Roberts, J. S.; Gruber, S. B.; Pienta, K. J.; Talpaz, M.; Chinnaiyan, A. M., Personalized oncology through integrative high–throughput sequencing: a pilot study. *Sci Transl Med* **2011,***3* (111), 111ra121.

20. Aboud, O. A.; Weiss, R. H., New opportunities from the cancer metabolome. *Clin Chem* **2013,***59* (1), 138–46.

21. Jiménez, B.; Mirnezami, R.; Kinross, J.; Cloarec, O.; Keun, H. C.; Holmes, E.; Goldin, R. D.; Ziprin, P.; Darzi, A.; Nicholson, J. K., ^{1}H HR–MAS NMR spectroscopy of tumor–induced local metabolic "field–effects" enables colorectal cancer staging and prognostication. *J Proteome Res* **2013,***12* (2), 959–68.

22. van der Greef, J.; van Wietmarschen, H.; van Ommen, B.; Verheij, E., Looking back into the future: 30 years of metabolomics at TNO. *Mass Spectrom Rev* **2013**.

23. Stella, C.; Beckwith–Hall, B.; Cloarec, O.; Holmes, E.; Lindon, J. C.; Powell, J.; van der Ouderaa, F.; Bingham, S.; Cross, A. J.; Nicholson, J. K., Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **2006,***5* (10), 2780–8.

24. Behrends, V.; Ebbels, T. M.; Williams, H. D.; Bundy, J. G., Time–resolved metabolic footprinting for nonlinear modeling of bacterial substrate utilization. *Appl Environ Microbiol* **2009,***75* (8), 2453–63.

25. Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X., Recent and potential developments of biofluid analyses in metabolomics. *J Proteomics* **2012,***75* (4), 1079–88.

26. Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X., Modern analytical techniques in metabolomics analysis. *Analyst* **2012,***137* (2), 293–300.

27. Dunn, W. B.; Ellis, D. I., Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry* **2005,***24* (4), 9.

28. Xiayan, L.; Legido–Quigley, C., Advances in separation science applied to metabonomics. *Electrophoresis* **2008,***29* (18), 3724–36.

29. Barbas, C.; Moraes, E. P.; Villaseñor, A., Capillary electrophoresis as a metabolomics tool for non–targeted fingerprinting of biological samples. *J Pharm Biomed Anal* **2011,***55* (4), 823–31.

30. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt–Kopplin, P., Liquid chromatography–mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **2013,***1292*, 51–65.

31. Álvarez–Sánchez, B.; Priego–Capote, F.; Luque de Castro, M. D., Metabolomics analysis I . Selection of biological samples and practical aspects preceding sample preparation. *Trends in Analytical Chemistry* **2010,***29* (2), 8.

32. Mamas, M.; Dunn, W. B.; Neyses, L.; Goodacre, R., The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch Toxicol* **2011,***85* (1), 5–17.

33. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry–based metabolomics. *Mass Spectrom Rev* **2007,***26* (1), 51–78.

34. Sumner, L. W.; Urbanczyk–Wochniak, E.; Broeckling, C. D., Metabolomics data analysis, visualization, and integration. *Methods Mol Biol* **2007,***406*, 409–36.

35. Lindon, J. C.; Nicholson, J. K., Analytical technologies for metabonomics and metabolomics, and multi–omic information recovery. *Trends in Analytical Chemistry* **2008,***27* (3), 194–204.

36. Nicholson, J. K.; Lindon, J. C., Systems biology: Metabonomics. *Nature* **2008,***455* (7216), 1054–6.

37. Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H. P., Targeted metabolomics for biomarker discovery. *Angew Chem Int Ed Engl* **2010,***49* (32), 5426–45.

38. Han, X.; Gross, R. W., Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *J Lipid Res* **2003,***44* (6), 1071–9.

39. Nicholson, J. K.; Holmes, E.; Wilson, I. D., Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* **2005,***3* (5), 431–8.

40. Lindon, J. C.; Nicholson, J. K., Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2008,***1*, 45–69.

41. Nicholson, J. K.; Holmes, E.; Kinross, J.; Burcelin, R.; Gibson, G.; Jia, W.; Pettersson, S., Host–gut microbiota metabolic interactions. *Science* **2012,***336* (6086), 1262–7.

42. Bictash, M.; Ebbels, T. M.; Chan, Q.; Loo, R. L.; Yap, I. K.; Brown, I. J.; de Iorio, M.; Daviglus, M. L.; Holmes, E.; Stamler, J.; Nicholson, J. K.; Elliott, P., Opening up the "Black Box": metabolic phenotyping and metabolome–wide association studies in epidemiology. *J Clin Epidemiol* **2010,***63* (9), 970–9.

43. Zhang, G. F.; Sadhukhan, S.; Tochtrop, G. P.; Brunengraber, H., Metabolomics, pathway regulation, and pathway discovery. *J Biol Chem* **2011,***286* (27), 23631–5.

44. Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C., Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal Chem* **2006,***78* (13), 4307–18.

45. Vuckovic, D., Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal Bioanal Chem* **2012,***403* (6), 1523–48.

46. Bollard, M. E.; Stanley, E. G.; Lindon, J. C.; Nicholson, J. K.; Holmes, E., NMR–based metabonomic approaches for evaluating physiological influences on biofluid composition. *NMR Biomed* **2005,***18* (3), 143–62.

47. Kuehnbaum, N. L.; Britz–McKibbin, P., New advances in separation science for metabolomics: resolving chemical diversity in a post–genomic era. *Chem Rev* **2013,***113* (4), 2437–68.

48. Issaq, H. J.; Van, Q. N.; Waybright, T. J.; Muschik, G. M.; Veenstra, T. D., Analytical and statistical approaches to metabolomics research. *J Sep Sci* **2009,***32* (13), 2183–99.

49. Nicholson, J. K.; Buckingham, M. J.; Sadler, P. J., High resolution 1H n.m.r. studies of vertebrate blood and plasma. *Biochem J* **1983,***211* (3), 605–15.

50. Theodoridis, G.; Gika, H. G.; Wilson, I. D., LC–MS–based methodology for global metabolite profiling in metabonomics/metabolomics. *Trends in Analytical Chemistry* **2008,***27* (3), 10.

51. Gamache, P. H.; Meyer, D. F.; Granger, M. C.; Acworth, I. N., Metabolomic applications of electrochemistry/mass spectrometry. *J Am Soc Mass Spectrom* **2004,***15* (12), 1717–26.

52. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2007,***2* (11), 2692–703.

53. Beckonert, O.; Coen, M.; Keun, H. C.; Wang, Y.; Ebbels, T. M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., High–resolution magic–angle–spinning NMR spectroscopy for metabolic profiling of intact tissues. *Nat Protoc* **2010,***5* (6), 1019–32.

54. Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W., LC–MS–based metabolomics. *Mol Biosyst* **2012,***8* (2), 470–81.

55. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC–MS. *Nat Protoc* **2010,***5* (6), 1005–18.

56. Wolfender, J. L.; Ndjoko, K.; Hostettmann, K., Liquid chromatography with ultraviolet absorbance–mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on–line structural investigation of plant metabolites. *J Chromatogr A* **2003,***1000* (1–2), 437–55.

57. Spagou, K.; Wilson, I. D.; Masson, P.; Theodoridis, G.; Raikos, N.; Coen, M.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Nicholson, J. K.; Want, E. J., HILIC–UPLC–MS for exploratory urinary metabolic profiling in toxicological studies. *Anal Chem* **2011,***83* (1), 382–90.

58. Guo, K.; Peng, J.; Zhou, R.; Li, L., Ion–pairing reversed–phase liquid chromatography fractionation in combination with isotope labeling reversed–phase liquid chromatography–mass spectrometry for comprehensive metabolome profiling. *J Chromatogr A* **2011,***1218* (23), 3689–94.

59. Donato, P.; Cacciola, F.; Tranchida, P. Q.; Dugo, P.; Mondello, L., Mass spectrometry detection in comprehensive liquid chromatography: basic concepts, instrumental aspects, applications and trends. *Mass Spectrom Rev* **2012,***31* (5), 523–59.

60. Garcia, A.; Barbas, C., Gas chromatography–mass spectrometry (GC–MS)–based metabolomics. *Methods Mol Biol* **2011,***708*, 191–204.

61. Mal, M.; Koh, P. K.; Cheah, P. Y.; Chan, E. C., Metabotyping of human colorectal cancer using two–dimensional gas chromatography mass spectrometry. *Anal Bioanal Chem* **2012,***403* (2), 483–93.

62. Ellis, D. I.; Goodacre, R., Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst* **2006,***131* (8), 875–85.

63. Milne, S. B.; Mathews, T. P.; Myers, D. S.; Ivanova, P. T.; Brown, H. A., Sum of the parts: mass spectrometry–based metabolomics. *Biochemistry* **2013,***52* (22), 3829–40.

64. Vaidyanathan, S.; Gaskell, S.; Goodacre, R., Matrix–suppressed laser desorption/ionisation mass spectrometry and its suitability for metabolome analyses. *Rapid Commun Mass Spectrom* **2006,***20* (8), 1192–8.

65. Aharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B., Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS* **2002,***6* (3), 217–34.

66. Ramautar, R.; Somsen, G.; de Jong, G., CE–MS in metabolomics. *Electrophoresis* **2009,***30* (1), 276–91.

67. Altamura, C.; Maes, M.; Dai, J.; Meltzer, H. Y., Plasma concentrations of excitatory amino acids, serine, glycine, taurine and histidine in major depression. *Eur Neuropsychopharmacol* **1995,***5 Suppl*, 71–5.

68. Hoekstra, R.; Fekkes, D.; Loonen, A. J.; Pepplinkhuizen, L.; Tuinier, S.; Verhoeven, W. M., Bipolar mania and plasma amino acids: increased levels of glycine. *Eur Neuropsychopharmacol* **2006,***16* (1), 71–7.

69. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J Proteome Res* **2007,***6* (2), 443–58.

70. Patel, K. N.; Patel, J. K.; Patel, M. P.; Rajput, G. C.; Patel, H. A., Introduction to hyphenated techniques and their applications in pharmacy. *Pharm Methods* **2010,***1* (1), 2–13.

71. Han, X.; Gross, R. W., Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom Rev* **2005,***24* (3), 367–412.

72. Quehenberger, O.; Armando, A. M.; Brown, A. H.; Milne, S. B.; Myers, D. S.; Merrill, A. H.; Bandyopadhyay, S.; Jones, K. N.; Kelly, S.; Shaner, R. L.; Sullards, C. M.; Wang, E.; Murphy, R. C.; Barkley, R. M.; Leiker, T. J.; Raetz, C. R.; Guan, Z.; Laird, G. M.; Six, D. A.; Russell, D. W.; McDonald, J. G.; Subramaniam, S.; Fahy, E.; Dennis, E. A., Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res* **2010,***51* (11), 3299–305.

73. Gross, R. W.; Han, X., Lipidomics at the interface of structure and function in systems biology. *Chem Biol* **2011,***18* (3), 284–91.

74. Issaq, H. J.; Abbott, E.; Veenstra, T. D., Utility of separation science in metabolomic studies. *J Sep Sci* **2008,***31* (11), 1936–47.

75. Ahmed, F. E., The role of capillary electrophoresis–mass spectrometry to proteome analysis and biomarker discovery. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,***877* (22), 1963–81.

76. Moraes, E. P.; Rupérez, F. J.; Plaza, M.; Herrero, M.; Barbas, C., Metabolomic assessment with CE–MS of the nutraceutical effect of Cystoseira spp extracts in an animal model. *Electrophoresis* **2011,***32* (15), 2055–62.

77. Goodwin, R. J.; Pitt, A. R.; Harrison, D.; Weidt, S. K.; Langridge–Smith, P. R.; Barrett, M. P.; Logan Mackay, C., Matrix–free mass spectrometric imaging using laser desorption ionisation Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* **2011,***25* (7), 969–72.

78. Kamleh, M. A.; Dow, J. A.; Watson, D. G., Applications of mass spectrometry in metabolomic studies of animal model and invertebrate systems. *Brief Funct Genomic Proteomic* **2009,***8* (1), 28–48.

79. Lao, Y. M.; Jiang, J. G.; Yan, L., Application of metabonomic analytical techniques in the modernization and toxicology research of traditional Chinese medicine. *Br J Pharmacol* **2009,***157* (7), 1128–41.

80. Lu, X.; Zhao, X.; Bai, C.; Zhao, C.; Lu, G.; Xu, G., LC–MS–based metabonomics analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,***866* (1–2), 64–76.

81. Almstetter, M. F.; Oefner, P. J.; Dettmer, K., Comprehensive two–dimensional gas chromatography in metabolomics. *Anal Bioanal Chem* **2012,***402* (6), 1993–2013.

82. Naz, S.; García, A.; Barbas, C., Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Anal Chem* **2013,***85* (22), 10941–8.

83. Keun, H. C.; Athersuch, T. J., Nuclear magnetic resonance (NMR)–based metabolomics. *Methods Mol Biol* **2011,***708*, 321–34.

84. Maher, A. D.; Fonville, J. M.; Coen, M.; Lindon, J. C.; Rae, C. D.; Nicholson, J. K., Statistical total correlation spectroscopy scaling for enhancement of metabolic information recovery in biological NMR spectra. *Anal Chem* **2012,***84* (2), 1083–91.

85. Katajamaa, M.; Oresic, M., Data processing for mass spectrometry–based metabolomics. *J Chromatogr A* **2007,***1158* (1–2), 318–28.

86. Tomasi, G.; van den Bergand, F.; Andersson, C., Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* **2004,***18*, 231–241.

87. Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K., Recursive segment–wise peak alignment of biological ^{1}H NMR spectra for improved metabolic biomarker recovery. *Anal Chem* **2009,***81* (1), 56–66.

88. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C., Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal Chem* **2006,***78* (7), 2262–7.

89. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Anal Chem* **2006,***78* (13), 4281–90.

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

91. García–Pérez, I.; Vallejo, M.; García, A.; Legido–Quigley, C.; Barbas, C., Metabolic fingerprinting with capillary electrophoresis. *J Chromatogr A* **2008,***1204* (2), 130–9.

92. Shockcor, J. P.; Holmes, E., Metabonomic applications in toxicity screening and disease diagnosis. *Curr Top Med Chem* **2002,***2* (1), 35–51.

93. Godzien, J.; Ciborowski, M.; Angulo, S.; Barbas, C., From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC–MS. *Electrophoresis* **2013,***34* (19), 2812–26.

94. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,***78* (3), 779–87.

95. Katajamaa, M.; Miettinen, J.; Oresic, M., MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **2006,***22* (5), 634–6.

96. Lommen, A., MetAlign: interface–driven, versatile metabolomics tool for hyphenated full–scan mass spectrometry data preprocessing. *Anal Chem* **2009,***81* (8), 3079–86.

97. Baran, R.; Kochi, H.; Saito, N.; Suematsu, M.; Soga, T.; Nishioka, T.; Robert, M.; Tomita, M., MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics* **2006,***7*, 530.

98. Jankevics, A.; Liepinsh, E.; Liepinsh, E.; Vilskersts, R.; Grinberga, S.; Pugovics, O.; Dambrova, M., Metabolomic studies of experimental diabetic urine samples by ^{1}H NMR spectroscopy and LC/MS method. *Chemometrics and Intelligent Laboratory Systems* **2009,***97*, 11–17.

99. Zhang, J.; Fang, A.; Wang, B.; Kim, S. H.; Bogdanov, B.; Zhou, Z.; McClain, C.; Zhang, X., iMatch: a retention index tool for analysis of gas chromatography–mass spectrometry data. *J Chromatogr A* **2011,***1218* (37), 6522–30.

100. Peetre, I. B., A Method and Program for Accurate Determination of Kováts' Retention Indices with Special Reference to Multicomponent Mixtures. *Chromatographia.* **1973,***6* (6), 9.

101. Zhang, J.; Koo, I.; Wang, B.; Gao, Q. W.; Zheng, C. H.; Zhang, X., A large scale test dataset to determine optimal retention index threshold based on three mass spectral similarity measures. *J Chromatogr A* **2012,***1251*, 188–93.

102. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot* **2005,***56* (410), 219–43.

103. Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier–Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Optimized preprocessing of ultra–performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem* **2011,***83* (15), 5864–72.

104. Want, E.; Masson, P., Processing and analysis of GC/LC–MS–based metabolomics data. *Methods Mol Biol* **2011,***708*, 277–98.

105. Chadeau–Hyam, M.; Campanella, G.; Jombart, T.; Bottolo, L.; Portengen, L.; Vineis, P.; Liquet, B.; Vermeulen, R. C., Deciphering the complex: methodological overview of statistical models to derive OMICS–based biomarkers. *Environ Mol Mutagen* **2013,***54* (7), 542–57.

106. Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J Proteome Res* **2007,***6* (2), 469–79.

107. Price, K. E.; Lunte, C. E.; Larive, C. K., Development of tissue–targeted metabonomics. Part 1. Analytical considerations. *J Pharm Biomed Anal* **2008,***46* (4), 737–47.

108. Rousseau, R.; Govaerts, B.; Verleysen, M.; Boulanger, B., Comparison of some chemometric tools for metabonomics biomarker identification. *Chemometrics and Intelligent Laboratory Systems.* **2008,***91* (1), 13.

109. Putri, S. P.; Yamamoto, S.; Tsugawa, H.; Fukusaki, E., Current metabolomics: technological advances. *J Biosci Bioeng* **2013,***116* (1), 9–16.

110. Westerhuis, J. A.; van Velzen, E. J.; Hoefsloot, H. C.; Smilde, A. K., Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics* **2010,***6* (1), 119–128.

111. Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., NMR–based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem Res Toxicol* **2008,***21* (1), 9–27.

112. van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,***7*, 142.

113. Chan, E. C.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography–mass spectrometry. *Nat Protoc* **2011,***6* (10), 1483–99.

114. Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; J. van Velzen, E. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A., Assessment of PLSDA cross validation. *Metabolomics.* **2008,***4* (1), 9.

115. *User Guide to SIMCA–P+. Version 12*. 2008; p 591.

116. Robinette, S. L.; Lindon, J. C.; Nicholson, J. K., Statistical spectroscopic tools for biomarker discovery and systems medicine. *Anal Chem* **2013,***85* (11), 5297–303.

117. Garcia–Perez, I.; Couto Alves, A.; Angulo, S.; Li, J. V.; Utzinger, J.; Ebbels, T. M.; Legido–Quigley, C.; Nicholson, J. K.; Holmes, E.; Barbas, C., Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating Schistosoma mansoni infection in a mouse model. *Anal Chem* **2010,***82* (1), 203–10.

118. Villaseñor, A.; Ramamoorthy, A.; dos Santos, M. S.; Lorenzo, M. P.; Laje, G.; Zarate, C.; Barbas, C.; Wainer, I. W., A pilot study of plasma metabolomic patterns from patients treated with ketamine for bipolar depression: Evidence for a response–related difference in mitochondrial networks. *British Journal of Pharmacology* **2013**, n/a–n/a.

119. Posma, J. M.; Robinette, S. L.; Holmes, E.; Nicholson, J. K., MetaboNetworks, an interactive Matlab–based toolbox for creating, customizing and exploring sub–networks from KEGG. *Bioinformatics* **2013**.

120. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M., KEGG for integration and interpretation of large–scale molecular data sets. *Nucleic Acids Res* **2012,***40* (Database issue), D109–14.

121. Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T., Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **2011,***7* (3), 307–328.

Journal of Pharmaceutical and Biomedical Analysis 55 (2011) 823–831

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Review

Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples

C. Barbas[a,*], Edgar P. Moraes[a,b], A. Villaseñor[a]

[a] CEMBIO (Center for Metabolomics and Bioanalysis), Facultad de Farmacia, Universidad San Pablo-CEU, Campus Montepríncipe, Boadilla del Monte, 28668 Madrid, Spain
[b] University of Sao Paulo, Brazil

ARTICLE INFO

Article history:
Received 7 October 2010
Received in revised form 31 January 2011
Accepted 1 February 2011
Available online 3 March 2011

Keywords:
Metabolomics
Metabonomics
Pattern recognition
Biomarkers

ABSTRACT

Metabolomics, understood as a data driven strategy trying to find markers of a situation under study without a priori hypothesis, has rapidly caught the attention and evolved from the simple pattern recognition strategy, which was a great innovation at its origins, to the interest for the final identification of markers responsible for class separation, i.e., from data to knowledge. Due to differences in physico-chemical properties and concentrations of the metabolites, but also due to differences in matrix properties, cross-platform approaches are proving to increase the capability of information. Once more techniques do not compete. This is the scene where capillary electrophoresis (CE) has its niche to provide information mainly on polar or ionic compounds in biological fluids.

General advantages and disadvantages of CE for sample fingerprinting will be discussed and methods will be classified depending on the detection system (UV or MS) as this strongly affects all the conditions. Recent developments will be presented in different biological fluids, although urine is without a doubt the preferred sample for CE analysis.

Contents

1. Introduction

The metabolome has been defined as the qualitative and quantitative collection of all low molecular weight molecules (metabolites) present in a cell that are participants in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell [1].

Metabolites include wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility, charge) properties. The degree of diversity includes ionic polar compounds, such as oxalate, low molecular weight (MW), volatile organic metabolites (urea), to the higher MW, polar (carbohydrates) and non-polar (terpenoids and lipids) metabolites. In addition, concentrations also extend over an estimated 7–9 orders of magnitude in concentration (pmol–mmol).

Metabolite analysis is not new; what is really new is the way we approach the problem. An omics technology has come to mean an approach capable of generating a comprehensive data

* Corresponding author. Tel.: +34 913 724711; fax: +34 913 510475.
E-mail address: cbarbas@ceu.es (C. Barbas).

0731-7085/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2011.02.001

set of whatever is being measured be it, genes (genomics), transcripts (transcriptomics), proteins (proteomics), or metabolites (metabol(n)omics). This is not only a new terminology, but also a real new approach to the analytical problem and with very different analytical requirements.

Thus with all metabolites as a goal in metabolomics, the technological obstacle seems to be the limiting step, being increasingly recognized that the combination of several techniques is currently the only possible approach.

Capillary electrophoresis has been considered a newcomer to the analytical tool box in metabolomics for years and honestly the adjective is becoming old-fashioned. CE is a technique that has not succeeded to get the main stream in analytical laboratories due to some drawbacks, but with a really high analytical potential for those who accomplish an expertise on it. Through this review we will critically discuss strengths and drawbacks of CE for metabolomics.

Classical analysis, both metabolite target analysis and metabolic profiling [2], is developed with the expectation that a measurement or combination of several measurements of related compounds reveals something about a specific problem or situation. For example, changes in oxalate and citrate reveal kidney problems. The analytical tools from sample treatment to the analytical method and adequate standards for calibration are specifically selected.

In an "omic" methodology the expectation is that the response pattern of numerous analytes, both known and unknown, is reflective of a situation and the comprehensive nature of the data set enables evaluation of systemic response. As an example, the urine fingerprint obtained with capillary electrophoresis of mice infected with schistosomiasis can be clearly differentiated from controls [3]. The end point to increase the knowledge about the pathology should be the identification of the specific biomarkers, being chemometrics a compulsory tool to approach data treatment.

Capillary electrophoresis in the simplest mode, capillary zone electrophoresis (CZE) is a separation technique based, on the differential transportation of charged species in an electric field through a conductive medium. Therefore, primary candidates for CE separation are ions. Since CE is based on a different separation mechanism to HPLC it also shows different selectivity and has been considered orthogonal and complementary to chromatography. Within CE systems, selectivity can be modified by changing the polarity, coating the capillaries or modifying the electrolyte. In MEKC, neutral solutes will partition with the micelles in a chromatographic way and selectivity differences can be caused by changing the nature and concentration of the surfactant. There is a huge bunch of possibilities, although some of them are limited by the detection technique employed, mainly in the case of mass spectrometry.

Main advantages of capillary electrophoresis compared with other analytical techniques used for biofluids fingerprinting:

(i) Small sample requirement (a few nanolitres), which is particularly well-suited when analysing valuable biological fluids through different analytical techniques.
(ii) Minor sample treatment. Just enough to avoid capillary clogging, because capillaries are rinsed after each run and apparently no irreversible retention is produced. That is good for a true metabolomic approach.
(iii) Ability to separate compounds in aqueous media. The main components of biological fluids display high polarity and water-solubility which makes them hard to retain on reversed-phase stationary phases. That makes it a suitable technique alternative to HPLC.
(iv) Capability of providing information complementary to mass, because analytes migration is based on mass to charge ratio.
(v) High efficiency which means high resolution
(vi) Multiple modes can be applied on the same sample to generate a broader picture what is closer to a real "omic" technique.

Main drawbacks

(i) Robustness is a weak point. That means in practice shifts in retention time which complicate serial analysis.
(ii) Sensitivity, because the volume of sample injected is in the nanoliters range and the cell pathway (the capillary diameter) is very narrow in UV. In MS commercially available interfaces require a sheath liquid which dilutes the analyte bands.

The preparation of samples for metabolomic analysis when a general fingerprint is the objective should be minimal to prevent any loss of metabolites. However, many important endogenous metabolites exist at very low concentrations in biological systems and the use of sample preconcentration techniques is often necessary to increase the sensitivity for the low-abundance endogenous metabolites. However, this topic is out of our focus.

The aim of this review is summarizing the state-of-art of capillary electrophoresis applied to biological sample fingerprinting. Due to the fact that the detection system is the component that most affect method development, data treatment and possible outcome of the measurements that will be the criterion for organizing the review.

There are several previous reviews with different coverage of the topic that can be complementary: Monton and Soga [4] described technical considerations and selected metabolomic applications of CE–MS; Ramautar et al. [5] demonstrated the applicability of CE in different fields of metabolomics research with illustrative clinical, bacterial and plant examples. The authors mainly focus on selected groups of metabolites, i.e. metabolic profiling, more than comprehensive fingerprinting, the review was later updated [6]. Barbas et al. reviewed the field of CE fingerprinting including four main blocks: sample treatment, analytical methods, chemometrics and applications heading focuses in urine, plasma, organic matter and plant extract studies [7]. Within a specific set of compounds, a recent review focuses on the analysis of nucleosides as metabolites of either oxidative DNA damage or RNA's turnover that are used as the potential tumor markers [8].

2. CE–UV

CE with UV detection is a really simple tool, when compared to more sophisticated ones such as ^{1}H NMR or mass spectrometry. Nevertheless, it has proved to be very useful for many applications and it would be an easy to transfer technique for field and routine analysis.

In a pioneering work Hanna-Brown et al. showed how sulphated b-CD-modified MEKC can be used to provide a useful tool for urine fingerprinting, allowing for separations of more than 80 signals in under 25 min [9]. In addition, Barbas and coworkers [10] showed that CE separations performed with reverse polarity, using polyacrylamide-coated capillaries, can provide the separation of over 60 signals in urine samples in around 25 min. They also proved that the possibility of measuring the same samples in CE operating with both two different polarities and buffer systems affords the exciting opportunity of having a truly holistic metabolomics approach, since all compounds in a given sample can then be resolved and detected in one or other mode, with the only limitation being UV sensitivity [11]. By using urine from control and diabetic rats, the complementarity of both CE methods, regarding the compounds detected in each one, was proved, and both were validated for a selected set of compounds with migration times across the total migration time window (for each mode). Com-

pounds fitting these criteria were, for the normal polarity CD–MEKC method: urea, creatinine, allantoin, hippuric acid and uric acid (as representative of cations and neutral compounds eluting first, followed by only large anions or weak acids at the end of the profile); and for the reverse polarity CZE method: oxalic, citric, benzoic and hippuric acids (showing mainly short chain organic acids from smaller to higher m/z).

2.1. Data pre-treatment in CE–UV fingerprints

Once obtained the omic fingerprint, before pattern recognition methods can be used for classification, preprocessing techniques for the denoising, baseline removal, normalisation and alignment of electropherograms must be applied.

CE in comparison to high performance liquid chromatography (HPLC) or gas chromatography (GC) produces less reproducible results, what may pose a problem in long lasting projects. Variations in migration time – a function of electroosmotic flow (EOF) inside the capillary, sample loading, wall interactions and physical errors (such as injection irreproducibility or temperature variations) – may lead to poorly reproducible data and preclude their appropriate interpretation [12] This is the most challenging point, while all the others share methodology with HPLC data pre-treatment and are well established.

Several methods have been proposed for signal alignment without prior knowledge of the compounds present; dynamic time warping (DTW) [13]; correlation optimised warping (COW) [14,15]; parametric time warping (PTW) [16,17]. The choice of a suitable method depends on the problem at hand. The DTW uses distance as a measure of similarity of two signals. However, in case of chromatographic signals, DTW is sensitive to different peak intensities. Therefore Nielsen proposed another approach, where the correlation coefficient is used as a similarity measure of two signals. PTW is based on a parametric model of warping functions that are created for calibration samples and then used in the correction of a whole batch of data. Different experiments have shown that the COW offers a better similarity measure [18,19].

Soria et al. [20] achieved alignment by combining the use of reference peaks with a method that uses information from multiple wavelengths to align electropherograms to a reference signal.

Bro et al. [21] suggested handling retention time shifts using the PARAFAC2 model. There is no need to align signals to equal length prior to data modeling, because the model does not assume parallel elution profiles. In this approach a covariance matrix of elution profiles preserving its 'inner-product structure' from sample to sample is used. PARAFAC2 is not a method for aligning as such, but it allows treating retention time shifts in chromatograms. It should be stressed that the method is suited for trilinear data, it means that it cannot be used for the univariate signals [22].

2.2. Data analysis

Data analysis is performed with common multivariate data analysis tools with each data point or peak area as a variable and using both non-supervised such as Principal Component Analysis (PCA) and supervised such as Partial Least Square Discriminant Analysis (PLS-DA) methods. There is no difference regarding data treatment with other metabolomic tools and the question is out of this review.

2.3. Selected applications

Urine fingerprints from control and diabetic rats have shown the clear effects of an antioxidant treatment on diabetic animals [23], which were not seen in controls, in a rapid, simple and cost-effective way without identifying a single marker. Classification was supported by that produced after PCA and PLS-DA of target variables obtained with selectively designed, time and reagent consuming methods [24]. Once developed, the methodology can be easily applied for a rapid in vivo screening of extracts with potential in vitro activity.

Among them, *Dunaliella salina* extracts were tested [25]. This is a microalga with high content in carotenoids that can be extracted via an environmentally clean process such as supercritical fluid extraction with CO_2. Five doses of *D. salina* extract with in vitro antioxidant properties were intragastrically administrated to adult male streptozotocin (STZ) diabetic rats. *D. salina* extract induced changes showed up by the multivariate analysis. Results of the treatment from most of the parameters can be considered beneficial for diabetic animals; although an increase in hyperglycemia and 8-isoprostane excretion when STZ treated animals received the extract was observed.

Schistosoma mansoni infection in mice has been fingerprinted using CE to study the capabilities of this technique as a diagnostic tool for this parasitic disease [3]. Results were compared with ultra high performance liquid chromatography–MS (UHPLC–MS) by applying the same set of multivariate analysis tools [26]. Principal component analysis of the aligned data provided a time trajectory where the infection was observed after 30 days with UPLC–MS and CE. Orthogonal partial least-square discriminant analysis models were validated for sensitivity and specificity. Multivariate data analysis derived from two different detection systems showed that CE–UV and UPLC–MS found equivalent results. In addition a validation was performed with samples from two sets of mice infected in an inter-laboratory experiment where different infection methods and animal husbandry procedures were employed in order to establish the core biological response to a *S. mansoni* infection. CE data were analysed using principal component analysis [27]. Validation of the scores consisted of permutation scrambling (100 repetitions) and a manual validation method, using a third of the samples (not included in the model) as a test or prediction set. The validation yielded 100% specificity and 100% sensitivity, demonstrating the robustness of these models with respect to deciphering metabolic perturbations in the mouse due to a *S. mansoni* infection. A total of 20 metabolites across the two experiments were identified that significantly discriminated between *S. mansoni*-infected and noninfected control samples.

Working with this model, chemometric tools have been applied to improve information extraction, which is probably the weak point of the technique: Matlab-based software employing the Pearson's correlation algorithm was applied to urine electropherograms from 20 mice infected with the schistosoma parasite [28]. The fingerprints were the sum of electropherograms analysed with normal and reverse polarity, in two different modes MEKC and CZE and with two different capillaries (uncoated and polyacrylamide coated) to provide a broad picture of the samples. Hippurate, a metabolite that was depleted in the infected group and is present in both polarities, was chosen as a test variable; it correlated with itself to a p value of ≤ 0.001. Phenylacetylglycine, was positively correlated to three metabolites in its same pathway. The study shows that the autocorrelation matrix is able to provide extra information from data files acquired by CE analyses.

Correlations were not only established within the same technique, but also with ^{1}H NMR. The two sets of metabolic profiles were first processed and analysed independently and were subsequently integrated using statistical correlation methods in order to facilitate cross assignment of metabolites. These correlation analyses facilitated structural elucidation using the identification power of one technique to enhance and validate the other, but also highlighted the enhanced ability to detect functional correlations between metabolites, thereby providing potential for achieving deeper mechanistic insight into the biological process [29].

In even more challenging assays CE–UV was used in human experiments where differences are not only due to disease or treatment but also genetic, environmental, nutritional and every kind of differences exists.

CE–UV was used to detect possible differences in urine of diabetic children as compared to controls [30]. Afterwards, children were enrolled in a trial to evaluate the effect of a special additive containing rosemary extract, vitamin E and PUFAs added to their standard diet through the meat to test a possible nutraceutical effect. Clear differences could be observed between treated and non-treated diabetic children and some of the metabolites associated could be identified. Fig. 1 shows the most representative metabolites identified in urine for each CZE in reverse polarity and MEKC in normal polarity.

CE was also employed to fingerprint bile collected at 10 min increments in the donors and in the recipients [31]. The electropherograms of these samples were aligned and normalised using correlation optimised warping algorithms and modelled with multivariate techniques. The resulting metabolic signatures were compared; in general donors and recipients showed distinct fingerprints and clustered separately. This study proposes CE fingerprinting of human bile as a promising technique to help unravel the complex metabolic pathways involved during transplantation.

Different strategies have been developed and successfully applied with CE–UV for profiling a specific set of compounds such as nucleosides, hormones and others in human biofluids, however the review will focus on non-target metabolite fingerprinting.

3. CE–MS

When coupling these two techniques the final result depends strongly on the characteristics of the interface system and type of mass spectrometer.

3.1. Interfaces and MS

Regarding the electrical interfacing, there are two types of instruments: (i) with the CE outlet capillary connected at ground and the electrospray voltage applied from the MS, which makes CE separation conditions independent from MS operation conditions, (ii) the end of the capillary is at a high voltage and it is reduced or increased by the electrospray voltage. In addition to the closure of the electrical circuit, a hydraulic flow rate is required to form the electrospray and two approaches have been developed, one the sheath flow interface, is commercially available and the other the sheathless interface has many publications, because dilution occurs at the mixing point, sensitivity tends to be compromised but is still under research. This problem can be avoided with a review by Zamfir [32] summarizes the state of the art. In any case, the optimization of a number of parameters: sheath liquid flow rate and composition, nebulizing gas, electrospray voltage, temperature, and some others are part of the method development in CE–MS.

One of the recent developments in interfaces for metabolomics [33] is related to a method that enables metabolomic fingerprinting of single cells and subcellular structures. A nebulizer-free coaxial sheath-flow interface completes the circuit and provides a stable electrospray, yielding a signal with a relative standard deviation of under 5% for the total ion electropherogram. Detection limits are in the low nanomolar range (i.e., 50 nM (300 amol)) for a number of cell-to-cell signaling molecules, including acetylcholine (ACh), histamine, dopamine, and serotonin. Single-cell electropherograms are reproducible, with a large number of metabolites detected; more than 100 compounds yield signals of over 104 counts from the injection of only 0.1% of the total content from a single metacerebral cell. Expected neurotransmitters are detected within the cells, as are compounds that have molecular masses consistent with all of the naturally occurring amino acids (except cysteine).

Regarding the ionization process, electrospray ionization (ESI) enables direct transfer of molecules from the liquid phase to the gas phase and it is in practice the only ion source employed in metabolomics.

Almost all types of mass analyzers have been coupled to CE. However, the typically narrow peaks resulting from CE separations, in addition to sample complexity require high mass accuracy and high resolution to be able to resolve closely migrating components with similar nominal masses. For that reason TOF are used in most of the applications.

Obviously the CE–MS coupling offers the structural identification capabilities of both mass spectrometer and migration time relationship to structure. Nevertheless, buffer systems and, therefore, separation modes have the typical constrains of MS, which are restricted to volatile buffers such as acetate and formiate. Recently Veuthey et al. [34] have reviewed fundamental concepts instrumental considerations and applications of the most widely used coupling in metabolomics: CE–TOF/MS.

3.2. Selected methods

Due to technical limitations both in CE migration and MS ionization a comprehensive analysis requires two or more sets of conditions, which have been optimised for different compound groups, and then concatenate the results. Amino acids, amines and nucleosides are analysed as cations using a very low pH electrolyte [35].

This was the case for Ramautar et al. [36] while studying urine of chronic patients with complex regional pain syndrome. Compounds with a higher number of amine groups in their structure, such as lysine or arginine, migrated first (10–11 min). Next a group of metabolites comprising amino acids (e.g., valine, leucine, threonine, pro line), vitamins (e.g., carnitine) and amines (e.g., creatine and N-methylnicotinamide) migrated together with small peptides like phenylalanine–glycine (12–14 min). Subsequently, the slower migrating amino acids comprising phenyl moieties (such as phenylalanine and tyrosine) migrated and then nucleosides like guanosine (14–17 min). Acidic metabolites, such as hippuric acid, are slightly anionic at pH 1.8 and, therefore, migrate just after the EOF. Similarly Minami et al. detected 953 peaks from blood plasma samples in mice. Among these peaks, 44 peaks exhibit significant circadian oscillations [37].

On the other hand, carboxylic acids, phosphorylated carboxylic acids, phosphorylated saccharides, nucleotides, and nicotinamide and flavinadenine coenzymes can be analysed as anions using an alkaline BGS and a cationic polymer-coated capillary (SMILE(+)) to reverse the EOF and prevent deleterious current drops [38]. For anions analysis Harada et al. [39] proposed a different approach based on pressure-assisted capillary electrophoresis coupled to electrospray ionization mass spectrometry. The feature of this method is that CE polarity is inverted from conventional CE analysis for anions, where the inlet of the capillary is at the anode and the outlet at the cathode. Moreover, an ordinary fused silica capillary was chosen instead of a specific cationic polymer-coated capillary (SMILE (+) capillary). A robust and inexpensive analytical method was established from the above-mentioned CE polarity and non-coated fused silica capillary. A trimethylamine acetate electrolyte, pH 10.0, provides a high resolution between isomers. Electrolyte flow using an air pump after electrophoresis enables the comprehensive and simultaneous analyses of sugar phosphates, organic acids, nucleotides and a CoA compound. According to the authors method can be used as a de facto standard for metabolome analysis.

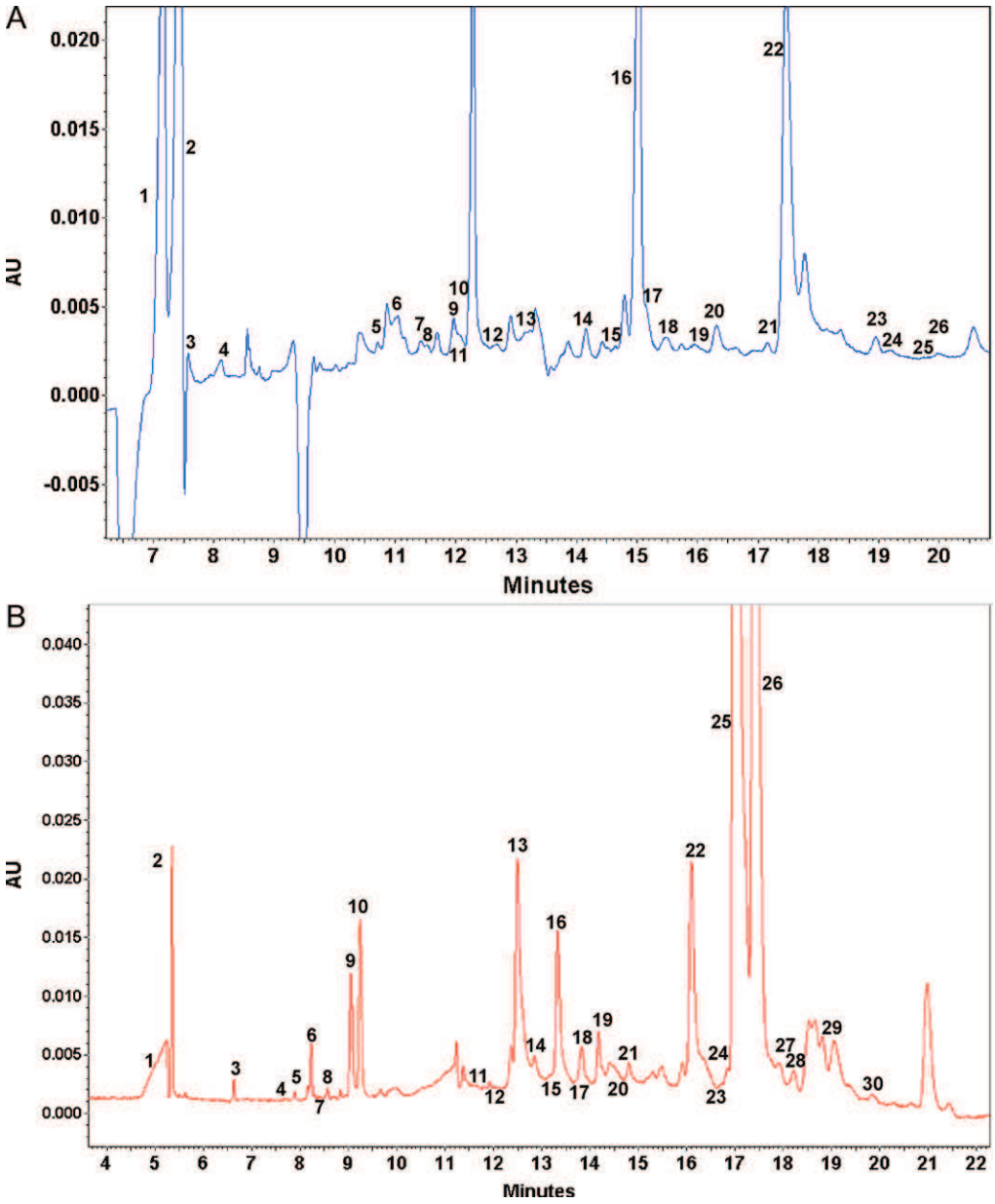


Fig. 1. Relevant metabolites identified in children urine samples by capillary electrophoresis for each CZE in reverse polarity and MEKC in normal polarity method. (A) Electropherogram profile for urinary sample obtained using the normal polarity CD–MEKC method. Peak identification: 1. Urea; 2. Creatinine; 3. Guanidinoacetate acid; 4. Ornithine; 5. Histidine; 6. Tyrosine, 7. Metionine; 8. Serine; 9. Phenylalanine; 10. 5-Methyluridine; 11. 2,5-Dihydroxybenzoate/glutamine; 12. Tryptophan; 13. Asparagine; 14. Phenaceturate; 15. Uridine; 16. Hippurate; 17. β-Phenyllactate; 18. Guanosine; 19. p-Hydroxyphenyllactate; 20. Quinaldate; 21. Inosine; 22. Urate; 23. Pyroglutamate; 24. β-3,4-Dihydroxyphenylalanine; 25. Isovanillate; 26. 3-Nitro-l-tyrosine. (B) Electropherogram profile for urinary sample obtained using the reverse polarity CZE method. Peak identification: 1. Chloride; 2. Nitrate; 3. Oxalate; 4. Fumarate; 5. Trans-aconitate; 6. 2-Ketoglutarate; 7. Malate/Succinate; 8. Isocitrate; 9. Glutarate; 10. Citrate; 11. Acetoacetate; 12. Lactate; 13. O-Phospho-L-Serine; 14. Glycerate; 15. 2-Hydroxybutyrate; 16. Benzoate; 17. 3-Hydroxybutyrate; 18. Pyroglutamate; 19. 2,5-Dihydroxybenzoate/glutamine; 20. 2,3-Dihydroxybenzoate; 21. Aspartate; 22. β-Phenyllactate; 23. Quinoline-2-carboxylate; 24. Glutamate; 25. Hippurate; 26. Urate; 27. p-Hydroxyphenyllactate; 28. Homovanillate; 29. Vanillylmandelate/2-aminoadipate; 30. Glucuronate (from [30]).

Only one paper has used pressurized capillary electrochromatography CEC (pCEC) with gradient elution for the metabolomic study of urinary samples [40] as an alternative chromatographic separation tool with higher degree of resolution, selectivity, sensitivity, and efficiency. The pCEC separation of urinary samples was performed on a RP column packed with C18, 5 μm particles with an ACN/water mobile phase containing TFA. The effects of the acid modifiers, applied voltage, mobile phase, and detection wavelength were systematically evaluated using eight spiked standards, as well as urine samples. A typical analytical trial of urine samples from Sprague Dawley rats exposed to high-energy diet was carried out following sample pretreatment. Significant differences in urinary metabolic profiles were observed between the high energy diet-induced obesity rats and the healthy control rats at the 6th week postdose. Multivariate statistical analysis revealed the differential metabolites in response to the diet, which were partially validated with the putative standards

Soga et al. [41] for the first time published the approach for the comprehensive and quantitative analysis of charged metabolites by capillary electrophoresis mass spectrometry with the two methods applied on the same samples. Their method enabled the determination of 352 metabolic standards and its utility was demonstrated in the analysis of 1692 metabolites from *Bacillus subtilis* extracts,

revealing significant changes in metabolites during *B. subtilis* sporulation.

Benavente et al. [42] also investigated a heterogeneous mixture of biologically relevant compounds in body fluids covering a broad range of physicochemical properties to optimise separation conditions in fused-silica capillaries. A running electrolyte containing 50 mM of acetic acid and 50 mM of formic acid at pH 2.5 was used for the CE separations. A sheath-flow electrospray interface was employed for CE–ESI–MS analysis. Sheath liquids containing 80:20 (v/v) methanol/water with 0.1% (v/v) of acetic acid or 60:40 (v/v) isopropanol/water with 0.5% (v/v) of ammonia were selected for optimum detection in positive and negative ESI modes, respectively.

The usefulness of noncovalently coated capillaries with layers of charged polymers was investigated by Ramautar et al. [43] to obtain global electrophoretic fingerprints of urinary metabolites covering a broad range of different compound classes in a highly repeatable way. Capillaries were coated with a bilayer of polybrene (PB) and poly(vinyl sulfonate) (PVS), or with a triple layer of PB, dextran sulfate (DS) and PB. The bilayer and triple layer coatings were evaluated at acidic (pH 2.0) and alkaline (pH 9.0) separation conditions, thereby providing separation conditions for basic and acidic compounds. A representative metabolite mixture and spiked urine samples were used for the evaluation of the four CE methods. Approximately, 600 molecular features were detected in rat urine by the PB–DS–PB CE–MS method whereas about 300 features were found with the PB–PVS CE–MS method. With regard to the analysis of anionic compounds by CE–MS, in general analyte responses were significantly lower than that for cationic compounds, most probably due to less efficient ionization and to ion suppression effects caused by the background electrolyte.

Lately Soga's group [44] optimised the method in the negative mode using a platinum electrospray ionization spray needle. They proved that the material in the spray needle had significant effect on the measurement of anions. A stainless steel spray needle was oxidized and corroded at the anodic electrode due to electrolysis. The use of a platinum ESI needle prevented both oxidation of the metals and needle corrosion. Sensitivity using the platinum needle increased from several- to 63-fold, with the largest improvements for anions exhibiting high metal chelating properties such as carboxylic acids, nucleotides, and coenzyme A compounds.

As a model study, human urine samples, obtained after intake of coffee, tea, or water, were analysed with capillary electrophoresis electrospray ionization time-of-flight mass spectrometry (CE–ESI–TOF-MS) [45]. In the model study highly significant differences due to beverage intake were obtained among the 10 first principal components.

A differential metabolomics strategy by capillary electrophoresis–electrospray ionization–mass spectrometry (CE–ESI–MS) to assess the efficacy of nutritional intervention to attenuate oxidative stress induced by strenuous exercise [46]. Time-dependent changes in global metabolism of filtered red blood cell lysates by CE–ESI–MS were measured to reveal a significant attenuation of cellular oxidation associated with high-dose oral N-acetyl-L-cysteine intake relative to a control. Untargeted metabolite profiling allowed for the identification and quantification of several putative early- and late-stage biomarkers that reflected oxidative stress inhibition due to nutritional intervention, including oxidized glutathione (GSSG), reduced glutathione (GSH), 3-methyl-histidine (3-MeHis), L-carnitine (CO), O-acetyl-L-carnitine (C2), and creatine (Cre).

Saliva is another readily accessible and informative biofluid, making it ideal for the early detection of a wide range of diseases including cardiovascular, renal, and autoimmune diseases, viral and bacterial infections and, importantly, cancers [47]. The authors conducted a comprehensive metabolite analysis of saliva samples obtained from 215 individuals (69 oral, 18 pancreatic and 30 breast cancer patients, 11 periodontal disease patients and 87 healthy controls) using capillary electrophoresis time-of-flight mass spectrometry (CE–TOF-MS). They identified 57 principal metabolites that can be used to accurately predict the probability of being affected by each individual disease.

Untargeted metabolomic analyses were performed by CE–ESI–MS on irradiated leukocytes exposed to increasing doses of radiation emitted from a Taylor source, which were subsequently incubated for 44 h to allow for cellular recovery [48]. CE–ESI–MS analysis of filtered white blood cell lysates was performed to quantify changes associated with 22 intra-cellular metabolites, which were consistently measured in leukocytes across all radiation levels.

Most cancer cells predominantly produce energy by glycolysis rather than oxidative phosphorylation via the tricarboxylic acid (TCA) cycle, even in the presence of an adequate oxygen supply (Warburg effect). However, little has been reported regarding the direct measurements of global metabolites in clinical tumor tissues. Capillary electrophoresis time-of-flight mass spectrometry was applied to simultaneously measure their levels in tumor and grossly normal tissues obtained from 16 colon and 12 stomach cancer patients [49]. Quantification of 94 metabolites in colon and 95 metabolites in stomach involved in glycolysis, the pentose phosphate pathway, the TCA and urea cycles, and amino acid and nucleotide metabolisms resulted in the identification of several cancer-specific metabolic traits. Extremely low glucose and high lactate and glycolytic intermediate concentrations were found in both colon and stomach tumor tissues, which indicated enhanced glycolysis and thus confirmed the Warburg effect. Significant accumulation of all amino acids except glutamine in the tumors implied autophagic degradation of proteins and active glutamine breakdown for energy production, i.e., glutaminolysis. In addition, significant organ-specific differences were found in the levels of TCA cycle intermediates, which reflected the dependency of each tissue on aerobic respiration according to oxygen availability.

3.3. Data treatment

Baseline correction, dataset normalisation and alignment, and the detection of significant differences between metabolite profiles through multivariate data analysis can be performed with the same tools as used for LC–MS. However, optimised software tools for CE–MS are needed [50] because of the sometimes large variation in migration time between runs and the wider variety of peak shapes in CE–MS data compared with LC–MS or GC–MS. Therefore, we implemented a stand-alone application named JDAMP (Java application for Differential Analysis of Metabolite Profiles), which allows users to identify the metabolites that vary between two groups. Soga's group has developed JDAMP with that purpose.

3.4. Metabolite identification

A major obstacle in metabolomics remains the identification and quantification of a large fraction of unknown metabolites in complex biological samples when purified standards are unavailable. Sugimoto et al. [51] have developed a method to identify unknown peaks based on the predicted migration time (tm) and accurate m/z values. They developed a predictive model using 375 standard cationic metabolites and support vector regression. The model yielded good correlations between the predicted and measured tm. Using the trained model, they subsequently predicted the tm for 2938 metabolites available from the public databases and assigned tentative identities to noise-filtered features in human urine samples. While 38.9% of the peaks were assigned metabolite names by

Table 1
CE applications for biofluid fingerprint in metabolomics.

Fluid	Pathology	Method	Markers	Reference
Urine, blood, liver and plasma	Diabetes	CD–MEKC 25 mM sodium tetraborate decahydrate, 75 mM SDS and 6.25 mM sulphated bcyclodextrin, pH 9.50, 2.25% (v/v) HFIP, 20 kV, 20 C. CZE polyacrylamide coated capillary, 0.2 M phosphoric acid, pH 6.10, 10% MeOH, −25 kV, 20 C		[25]
Red blood cell	Theoretical work	CE–ESI–MS electrolite 1.4 M formic acid, pH 1.8, sheath liquid 1:1 MeOH/H_2O with 0.1% FA, 25 kV, 20 C	Amino acids, amines, peptides, acylcarnitines, nucleosides	[52]
Urine	Schistosoma	MECK 25 mM sodium borate, 75 mM SDS and 6.25 mM sulphated b-cyclodextrin, pH 9.5, 18 kV, 20 C and CZE 0.2 M phosphate, 10% MeOH, pH 6.2	Hippurate and phenylacetyl-glycine	[28]
Tumor tissues	Colon and stomach cancer	CE–TOFMS cationic – BGE 1 M formic acid. SL 1:1 H_2O/MeOH, 0.5 μM reserpine, 30 kV, 20 C CE–TOFMS anionic, cap cationic polymer-coated SMILE (+) – BGE 50 mM ammonium acetate, pH 8.5, SL ammonium acetate (5 mM), 1:1 MeOH/H_2O, 1 μM reserpine, −30 kV	95 metabolites	[49]
Neurons	Study of neural functions	CE–ESI–MS electrolite 1% formic acid, sheath liquid 1:1 MeOH/H_2O with 0.1% FA, 20 kV	Neurotransmitters including acetylcholine (ACh), histamine, dopamine, and serotonin	[33]
Mouse liver	Oxidative stress	CE–TOFMS negative mode-coated with a cationic polymer BGE 50 mM acetic acid (pH 3.4), SL 5 mM ammonium acetate in 50% (v/v) methanol–water containing 0.1 μM Hexakis, −30 kV, 20 C	Anionic metabolites	[44]
Red blood cell	Oxidative stress	CE–ESI–MS cationic mode – BGE 1.0 M formic acid, pH 1.8, SL 0.1% formic acid in 1:1 MeOH/H_2O – negative mode – BGE 50 mM ammonium acetate pH 8.5, SL 5 mM ammonium acetate in 1:1 MeOH/H_2O, 30 kV, 20 C	Glutathione	[54]
Urine	Diabetes	CE–UV	Antioxidants	[23]
Urine	Urinary tract infection	CE–TOFMS capillaries noncovalently coated, electrolite 1 M formic acid, pH 1.8, sheath liquid 50/50 H_2O/MeOH 0,1% FA, 30 kV, 25 C	Amino acids	[55]
Urine	Steroid hormone metabolism	MEKC–ESI–MS partial filling	Steroids	[57]
Plasma, urine and cerebrospinal fluid	Complex regional pain syndrome	PB–DS–PB coating: BGE, 200 mM sodium phosphate (pH 6.0); injection, 35 mbar for 10 s; detection, UV at 200 nm, −10 kV, 25 C – PB–PVS coating: BGE, 1 M Formic acid (pH 1.8); sample injection, 90 mbar for 90 s; preinjection, ammonium hydroxide (12.5%) at 50 mbar for 9 s; 30 kV, 25 C	Organic acids and amino acids	[58]
Saliva	Oxidative stress	Microchip 30 mM borate buffer, pH 9.5, 30 kV	DHBA	[56]
Urine	Diabetes	CD–MEKC and CZE	Antioxidants	[11]
Urine	*Schistosoma mansoni* infection	EKC 25 mM sodium borate, 75 mM SDS and 6.25 mM sulfated b-CD, pH 9,5, 18 kV, 20 C. CZE coated capillary 0.2 M phosphoric acid, 10% MeOH, pH 6,2, −20 kV, 25 C	Many	[3]
Human bile	Liver transplantation	10 mM sodium borate, 10 mM sodium dihydrogen, phosphate adjusted to pH 7 with diluted phosphoric acid, to which 50 mM SDS, 5 mM b-CD and 5 mM HP-b-CD, 10% (v/v) CAN, 20% MeOH, 25 kV, 30 C	Bile acids	[31]
Rat urine	New method to improve positive charge in acids	CE–ESI–MS BGE 1 M FA. SL 50/50 H_2O/MeOH 0.1% FA, 30 kV	Carboxylic acids	[59]
E. coli	Strategy to find low-abundance metabolites	CE–ESI–ITMS BGE 1 M FA. SL 1:1 H_2O/MeOH, 0,1% FA. 23 kV, 20 C	Amino acids	[60]
Urine	Obese rats	CEC with gradient elution		[40]
Human red-cell	Hypoxia	Mathematical model compared with CE–MS	Metabolites of glycolysis	[61]
Human cerebrospinal fluid	Neurobiological disorders	CE–ESI–TOFMS – modification of the silica-wall with 1-(4-iodobutyl)4-aza-1-azoniabicyclo[2,2,2]octane iodide, also named M7C4I, has successfully been used to deactivate the fused silica wall and generate a stable reversed electroosmotic flow	Kynurenic acid	[62]
Mouse liver	Acetaminophen-induced hepatotoxicity	CE–TOFMS cationic – BGE 1 M formic acid. SL 1:1 H_2O/MeOH, 0.5 M reserpine, 30 kV, 20 C CE–TOFMS anionic, cap cationic polymer-coated SMILE (+) – BGE 50 mM ammonium acetate, pH 8.5, SL ammonium acetate (5 mM), 1:1 MeOH/H_2O, 20 M PIPES, 1 M reserpine, −30 kV, 20 C	Ophthalmic acid	[63]
E. coli	Discover novel enzymatic activities	–	Sugars/sugars phosphates	[64]
Urine	Study of urine stability	Sodium borate/SDS/sulphated b-cyclodextrin (25/75/6.25 mM), pH 9.50; 18 kV, 20 C, 5 s (0.5 psi); 200 nm	Urea, creatinine, phenylalanine, hippuric acid, uric acid	[65]

Table 1 (*Continued*)

Fluid	Pathology	Method	Markers	Reference
In vitro assay	Approach designed to reconstitute a metabolic pathway	CE–ESI–MS cationic polymer (polybrene)-coated capillary – BGE 50 mM ammonium acetate, pH 8.5 – SL 5 mM ammonium acetate in 50% (v/v) methanol–water, −30 kV, 20 C	Glycolysis metabolites	[67]
E. coli	C-13-labeling experiment	CE–TOF-MS	Amino acids	[66]
Bacilus subtilis extracts	Changes during *B. subtilis* sporulation		1692 metabolites	[41]
Bacilus subtilis extracts	Comprehensive analysis of metabolic intermediates	CE–ESI–MS cationic polymer (polybrene)-coated capillary – BGE 50 mM ammonium acetate, pH 9.0	32 carboxylic acids	[68]

matching with the standard library alone, the proportion increased to 52.2% with this method. However, only the description of the procedure, but not the list of metabolites is publicly available.

Kenneth et al. [52] introduced a multivariate strategy for de novo quantification of cationic/zwitterionic metabolites using capillary electrophoresis–electrospray ionization–mass spectrometry (CE–ESI–MS) based on fundamental molecular, thermodynamic, and electrokinetic properties of an ion. The applicability of the multivariate model to quantify micromolar levels of metabolites spiked in red blood cell lysates was also examined by CE–ESI–MS without significant matrix effects caused by involatile salts and/or major co-ion interferences.

To conclude, it is necessary to comment that there are some publications comparing different analytical techniques. Büscher et al. [53] developed a systematic cross-platform comparison of different separation and detection methods for quantitative metabolomics. Using a defined mixture of 91 metabolites (covering glycolysis, pentose phosphate pathway, the tricarboxylic acid (TCA) cycle, redox metabolism, amino acids, and nucleotides), they compared six separation methods designed for the analysis of these mostly very polar primary metabolites, two methods each for gas chromatography (GC), liquid chromatography (LC), and capillary electrophoresis (CE). According to their conclusions, for analyses on a single platform, LC provides the best combination of both versatility and robustness. If a second platform can be used, it is best complemented by GC. In our experience, interesting information can be obtained from this type of synthetic mixtures, however, they do not reproduce the real problems in complex biological matrices and it is there where capillary electrophoresis can find its main application. Urine contains an important amount of polar or ionic compounds poorly retained in reversed phase LC. The aqueous sample should be diluted in a higher proportion of organic solvent to analyse in HILIC mode, however, frequently urine is preserved with sodium borate and HILIC mode does not work properly in the presence of salts. In addition, in GC it produces a huge peak of urea that precludes the observation of most of the compounds in the profile. When uricase is employed changes in a number of metabolites have been described. A similar situation happens with cell culture media which contain mainly amino acids, salts and glucose. For that reason CE can be a good complementary separation technique in metabolomics for polar and ionic compounds in biological samples.

Main applications in CE for biological sample fingerprinting are summarized in Table 1 for an easy overview.

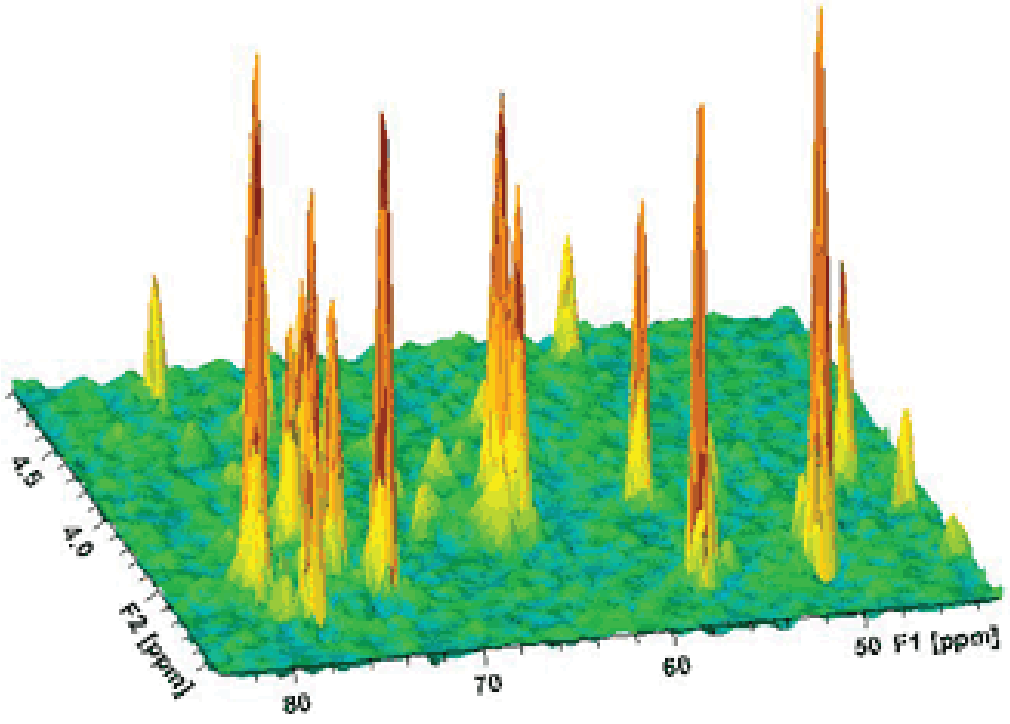
Acknowledgements

The authors gratefully acknowledge the financial support from Ministery of Science and Technology (MCIT) CTQ2008-03779 and Comunidad de Madrid, S-GEN-0247-2006.

References

[1] Metabolic Profiling: its Role in Biomarker Discovery and Gene Function Analysis, Kluwer Academic Publishers, London, UK, 2003.
[2] O. Fiehn, J. Kopka, P. Dörmann, T. Altmann, R. Trethewey, L. Willmitzer, Metabolite profiling for plant functional genomics, Nat. Biotechnol. 18 (2000) 1157–1161.
[3] I. García-Pérez, P. Whitfield, A. Bartlett, S. Angulo, C. Legido-Quigley, M. Hanna-Brown, C. Barbas, Metabolic fingerprinting of *Schistosoma mansoni* infection in mice urine with capillary electrophoresis, Electrophoresis 29 (2008) 3201–3206.
[4] M. Monton, T. Soga, Metabolome analysis by capillary electrophoresis-mass spectrometry, J. Chromatogr. A 1168 (2007) 237–246.
[5] R. Ramautar, A. Demirci, G.J. de Jong, Capillary electrophoresis in metabolomics, Trends Anal. Chem. 25 (2006) 455–466.
[6] R. Ramautar, G. Somsen, G. de Jong, CE–MS in metabolomics, Electrophoresis 30 (2009) 276–291.
[7] I. García-Pérez, M. Vallejo, A. García, C. Legido-Quigley, C. Barbas, Metabolic fingerprinting with capillary electrophoresis, J. Chromatogr. A 1204 (2008) 130–139.
[8] M.J. Markuszewski, W. Struck, M. Waszczuk-Jankowska, R. Kaliszan, Metabolomic approach for determination of urinary nucleosides as potential tumor markers using electromigration techniques, Electrophoresis 31 (2010) 2300–2310.
[9] C. Guillo, D. Barlow, D. Perrett, M. Hanna-Brown, Micellar electrokinetic capillary chromatography and data alignment analysis: a new tool in urine profiling, J. Chromatogr. A 1027 (2004) 203–212.
[10] A. García, C. Barbas, R. Aguilar, M. Castro, Capillary electrophoresis for rapid profiling of organic acidurias, Clin. Chem. 44 (1998) 1905–1911.
[11] C. Barbas, M. Vallejo, A. García, D. Barlow, M. Hanna-Brown, Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies, J. Pharm. Biomed. Anal. 47 (2008) 388–398.
[12] B. Casado, C. Zanone, L. Annovazzi, P. Iadarola, G. Whalen, J. Baraniuk, Urinary electrophoretic profiles from chronic fatigue syndrome and chronic fatigue syndrome/fibromyalgia patients: a pilot study for achieving their normalization, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 814 (2005) 43–51.
[13] A. Kassidas, J.F. MacGregor, P.A. Taylor, Synchronization of batch trajectories using dynamic time warping, AIChE J. 44 (1998) 864–875.
[14] N.-P.V. Nielsena, J.M. Carstensenb, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr. A 805 (1998) 17–35.
[15] G. Tomasi, F. van den Bergand, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241.
[16] A. van Nederkassel, M. Daszykowski, P. Eilers, Y. Heyden, A comparison of three algorithms for chromatograms alignment, J. Chromatogr. A 1118 (2006) 199–210.
[17] P. Eilers, Parametric time warping, Anal. Chem. 76 (2004) 404–411.
[18] E. Szymańska, M. Markuszewski, X. Capron, A. van Nederkassel, Y. Vander Heyden, M. Markuszewski, K. Krajka, R. Kaliszan, Evaluation of different warping methods for the analysis of CE profiles of urinary nucleosides, Electrophoresis 28 (2007) 2861–2873.
[19] E. Szymańska, M. Markuszewski, X. Capron, A. van Nederkassel, Y. Heyden, M. Markuszewski, K. Krajka, R. Kaliszan, Increasing conclusiveness of metabonomic studies by chem-informatic preprocessing of capillary electrophoretic data on urinary nucleoside profiles, J. Pharm. Biomed. Anal. 43 (2007) 413–420.
[20] A. Soria, B. Wright, D. Goodall, J. Wilson, Data processing in metabolic fingerprinting by CE–UV: application to urine samples from autistic children, Electrophoresis 28 (2007) 950–964.
[21] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2: Part II. Modeling chromatographic data with retention time shifts, J. Chemom. 13 (1999) 295–309.
[22] V. Pravdova, B. Walczak, D.L. Massart, A comparison of two algorithms for warping of analytical signals, Anal. Chim. Acta 456 (2002) 77–92.
[23] M. Vallejo, S. Angulo, D. García-Martínez, A. García, C. Barbas, New perspective of diabetes response to an antioxidant treatment through metabolic fingerprinting of urine by capillary electrophoresis, J. Chromatogr. A 1187 (2008) 267–274.
[24] F. Rupérez, D. García-Martínez, B. Baena, N. Maeso, A. Cifuentes, C. Barbas, E. Herrera, Evolution of oxidative stress parameters and response to oral vitamins E and C in streptozotocin-induced diabetic rats, J. Pharm. Pharmacol. 60 (2008) 871–878.
[25] F. Ruperez, D. Garcia-Martinez, B. Baena, N. Maeso, M. Vallejo, S. Angulo, A. Garcia, E. Ibañez, F. Señorans, A. Cifuentes, C. Barbas, *Dunaliella salina* extract effect on diabetic rats: metabolic fingerprinting and target metabolite analysis, J. Pharm. Biomed. Anal. 49 (2009) 786–792.

[26] I. Garcia-Perez, M. Earll, S. Angulo, C. Barbas, C. Legido-Quigley, Chemometric analysis of urine fingerprints acquired by liquid chromatography–mass spectrometry and capillary electrophoresis: application to the schistosomiasis mouse model, Electrophoresis 31 (2010) 2349–2355.
[27] I. Garcia-Perez, S. Angulo, J. Utzinger, E. Holmes, C. Legido-Quigley, C. Barbas, Chemometric and biological validation of a capillary electrophoresis metabolomic experiment of *Schistosoma mansoni* infection in mice, Electrophoresis 31 (2010) 2338–2348.
[28] S. Angulo, I. García-Pérez, C. Legido-Quigley, C. Barbas, The autocorrelation matrix probing biochemical relationships after metabolic fingerprinting with CE, Electrophoresis 30 (2009) 1221–1227.
[29] I. Garcia-Perez, A. Couto Alves, S. Angulo, J. Li, J. Utzinger, T. Ebbels, C. Legido-Quigley, J. Nicholson, E. Holmes, C. Barbas, Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating *Schistosoma mansoni* infection in a mouse model, Anal. Chem. 82 (2010) 203–210.
[30] C. Balderas, A. Villaseñor, A. García, F. Rupérez, E. Ibañez, J. Señorans, J. Guerrero-Fernández, I. González-Casado, R. Gracia-Bouthelier, C. Barbas, Metabolomic approach to the nutraceutical effect of rosemary extract plus omega-3 PUFAs in diabetic children with capillary electrophoresis, J. Pharm. Biomed. Anal. 53 (2010) 1298–1304.
[31] K. Papaspyridonos, I. Garcia-Perez, S. Angulo, P. Domann, H. Vilca-Melendez, N. Heaton, G. Murphy, E. Holmes, C. Barbas, C. Legido-Quigley, Fingerprinting of human bile during liver transplantation by capillary electrophoresis, J. Sep. Sci. 31 (2008) 3058–3064.
[32] A. Zamfir, Recent advances in sheathless interfacing of capillary electrophoresis and electrospray ionization mass spectrometry, J. Chromatogr. A 1159 (2007) 2–13.
[33] T. Lapainis, S. Rubakhin, J. Sweedler, Capillary electrophoresis with electrospray ionization mass spectrometric detection for single-cell metabolomics, Anal. Chem. 81 (2009) 5858–5864.
[34] A. Staub, J. Schappler, S. Rudaz, J. Veuthey, CE–TOF/MS: fundamental concepts, instrumental considerations and applications, Electrophoresis 30 (2009) 1610–1623.
[35] T. Soga, D. Heiger, Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry, Anal. Chem. 72 (2000) 1236–1241.
[36] R. Ramautar, A. van der Plas, E. Nevedomskaya, R. Derks, G. Somsen, G. de Jong, J. van Hilten, A. Deelder, O. Mayboroda, Explorative analysis of urine by capillary electrophoresis–mass spectrometry in chronic patients with complex regional pain syndrome, J. Proteome Res. 8 (2009) 5559–5567.
[37] Y. Minami, T. Kasukawa, Y. Kakazu, M. Iigo, M. Sugimoto, S. Ikeda, A. Yasui, G. van der Horst, T. Soga, H. Ueda, Measurement of internal body time by blood metabolomics, Proc. Natl. Acad. Sci. U.S.A 106 (2009) 9890–9895.
[38] T. Soga, Y. Ueno, H. Naraoka, K. Matsuda, M. Tomita, T. Nishioka, Pressure-assisted capillary electrophoresis electrospray ionization mass spectrometry for analysis of multivalent anions, Anal. Chem. 74 (2002) 6224–6229.
[39] K. Harada, E. Fukusaki, A. Kobayashi, Pressure-assisted capillary electrophoresis mass spectrometry using combination of polarity reversion and electroosmotic flow for metabolomics anion analysis, J. Biosci. Bioeng. 101 (2006) 403–409.
[40] G. Xie, M. Su, P. Li, X. Gu, C. Yan, Y. Qiu, H. Li, W. Jia, Analysis of urinary metabolites for metabolomic study by pressurized CEC, Electrophoresis 28 (2007) 4459–4468.
[41] T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, T. Nishioka, Quantitative metabolome analysis using capillary electrophoresis mass spectrometry, J. Proteome Res. 2 (2003) 488–494.
[42] F. Benavente, R. van der Heijden, U. Tjaden, J. van der Greef, T. Hankemeier, Metabolite profiling of human urine by CE–ESI–MS using separation electrolytes at low pH, Electrophoresis 27 (2006) 4570–4584.
[43] R. Ramautar, J. Toraño, G. Somsen, G. de Jong, Evaluation of CE methods for global metabolic profiling of urine, Electrophoresis 31 (2010) 2319–2327.
[44] T. Soga, K. Igarashi, C. Ito, K. Mizobuchi, H. Zimmermann, M. Tomita, Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry, Anal. Chem. 81 (2009) 6165–6174.
[45] E. Allard, D. Bäckström, R. Danielsson, P. 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.
[46] R. Lee, D. West, S. Phillips, P. Britz-McKibbin, Differential metabolomics for quantitative assessment of oxidative stress with strenuous exercise and nutritional intervention: thiol-specific regulation of cellular metabolism with N-acetyl-L-cysteine pretreatment, Anal. Chem. 82 (2010) 2959–2968.
[47] M. Sugimoto, D. Wong, A. Hirayama, T. Soga, M. Tomita, Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles, Metabolomics 6 (2010) 78–95.
[48] R. Lee, P. Britz-McKibbin, Metabolomic studies of radiation-induced apoptosis of human leukocytes by capillary electrophoresis–mass spectrometry and flow cytometry: adaptive cellular responses to ionizing radiation, Electrophoresis 31 (2010) 2328–2337.
[49] A. Hirayama, K. Kami, M. Sugimoto, M. Sugawara, N. Toki, H. Onozuka, T. Kinoshita, N. Saito, A. Ochiai, M. Tomita, H. Esumi, T. Soga, Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry, Cancer Res 69 (2009) 4918–4925.
[50] M. Sugimoto, A. Hirayama, T. Ishikawa, M. Robert, R. Baran, K. Uehara, K. Kawai, T. Soga, M. Tomita, Differential metabolomics software for capillary electrophoresis–mass spectrometry data analysis, Metabolomics 6 (2009) 27–41.
[51] M. Sugimoto, A. Hirayama, M. Robert, S. Abe, T. Soga, M. Tomita, Prediction of metabolite identity from accurate mass, migration time prediction and isotopic pattern information in CE–TOFMS data, Electrophoresis 31 (2010) 2311–2318.
[52] K. Chalcraft, R. Lee, C. Mills, P. Britz-McKibbin, Virtual quantification of metabolites by capillary electrophoresis–electrospray ionization–mass spectrometry: predicting ionization efficiency without chemical standards, Anal. Chem. 81 (2009) 2506–2515.
[53] J. Büscher, D. Czernik, J. Ewald, U. Sauer, N. Zamboni, Cross-platform comparison of methods for quantitative metabolomics of primary metabolism, Anal. Chem. 81 (2009) 2135–2143.
[54] R. Lee, P. Britz-McKibbin, Differential rates of glutathione oxidation for assessment of cellular redox status and antioxidant capacity by capillary electrophoresis–mass spectrometry: an elusive biomarker of oxidative stress, Anal. Chem. 81 (2009) 7047–7056.
[55] R. Ramautar, O.A. Mayboroda, R.J. Derks, C. van Nieuwkoop, J.T. van Dissel, G.W. Somsen, A.M. Deelder, G.J. de Jong, Capillary electrophoresis–time of flight-mass spectrometry using noncovalently bilayer-coated capillaries for the analysis of amino acids in human urine, Electrophoresis 29 (2008) 2714–2722.
[56] D.J. Marchiarullo, J.Y. Lim, Z. Vaksman, J.P. Ferrance, L. Putcha, J.P. Landers, Towards an integrated microfluidic device for spaceflight clinical diagnostics microchip-based solid-phase extraction of hydroxyl radical markers, J. Chromatogr. A 1200 (2008) 198–203.
[57] H. Sirén, T. Seppänen-Laakso, M. Oresic, Capillary electrophoresis with UV detection and mass spectrometry in method development for profiling metabolites of steroid hormone metabolism, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 375–382.
[58] R. Ramautar, O. Mayboroda, A. Deelder, G. Somsen, G. de Jong, Metabolic analysis of body fluids by capillary electrophoresis using noncovalently coated capillaries, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 370–374.
[59] W.C. Yang, F.E. Regnier, J. Adamec, Comparative metabolite profiling of carboxylic acids in rat urine by CE–ESI MS/MS through positively pre-charged and (2)H-coded derivatization, Electrophoresis 29 (2008) 4549–4560.
[60] R. Lee, A.S. Ptolemy, L. Niewczas, P. Britz-McKibbin, Integrative metabolomics for characterizing unknown low-abundance metabolites by capillary electrophoresis–mass spectrometry with computer simulations, Anal. Chem. 79 (2007) 403–415.
[61] A. Kinoshita, K. Tsukada, T. Soga, T. Hishiki, Y. Ueno, Y. Nakayama, M. Tomita, M. Suematsu, Roles of hemoglobin allostery in hypoxia-induced metabolic alterations in erythrocytes: simulation and its verification by metabolome analysis, J. Biol. Chem. 282 (2007) 10731–10741.
[62] B. Arvidsson, N. Johannesson, A. Citterio, P.G. Righetti, J. Bergquist, High throughput analysis of tryptophan metabolites in a complex matrix using capillary electrophoresis coupled to time-of-flight mass spectrometry, J. Chromatogr. A 1159 (2007) 154–158.
[63] T. Soga, R. Baran, M. Suematsu, Y. Ueno, S. Ikeda, T. Sakurakawa, Y. Kakazu, T. Ishikawa, M. Robert, T. Nishioka, M. Tomita, Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption, J. Biol. Chem. 281 (2006) 16768–16776.
[64] N. Saito, M. Robert, S. Kitamura, R. Baran, T. Soga, H. Mori, T. Nishioka, M. Tomita, Metabolomics approach for enzyme discovery, J. Proteome Res. 5 (2006) 1979–1987.
[65] C. Guillo, D. Perrett, M. Hanna-Brown, Validation and further optimisation of a cyclodextrin-modified micellar electrokinetic capillary chromatography method for urine profiling, Chromatogr. Suppl. 59 (2004) S157–S164.
[66] Y. Toya, N. Ishii, T. Hirasawa, M. Naba, K. Hirai, K. Sugawara, S. Igarashi, K. Shimizu, M. Tomita, T. Soga, Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis, J. Chromatogr. A 1159 (2007) 134–141.
[67] A. Itoh, Y. Ohashi, T. Soga, H. Mori, T. Nishioka, M. Tomita, Application of capillary electrophoresis–mass spectrometry to synthetic in vitro glycolysis studies, Electrophoresis 25 (2004) 1996–2002.
[68] T. Soga, Y. Ueno, H. Naraoka, Y. Ohashi, M. Tomita, T. Nishioka, Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry, Anal. Chem. 74 (2002) 2233–2239.



FIRST SECTION.

METHOD DEVELOPMENT AND VALIDATION.

CHAPTER 2

1. Gas Chromatography–Mass Spectrometry (GC–MS) and Liquid Chromatography–Mass Spectrometry (LC–MS)

This Chapter will be fully dedicated to the introduction of Gas Chromatography–Mass Spectrometry (GC–MS) while Liquid Chromatography–Mass Spectrometry (LC–MS) will be described in depth in Chapter 5, where an application using this technique only is shown.

2.1 Fundamentals of Gas Chromatography (GC)

Chromatography was invented in the early 1900s by M. S. Tswett[1]. It was defined as a separation method aimed to resolve the components in a mixture, on the basis of consecutive equilibria that results from the interaction of the compounds between two phases: the first fixed or stationary, and the second moving in contact with the first phase[2]. When the mobile phase is a gas this is defined as "*gas chromatography (GC)*". In GC, the mobile phase should be an inert, dry and pure gas, i.e. helium, argon, or nitrogen, which functions as a carrier, taking the compounds through the column to the detector. For GC–MS, helium is preferable over others. The stationary phases are generally set on open tubular columns that are known as "*capillary columns*"; these can have different inner diameters (0.1 to 0.25mm), can be up to 30m in length, and allow for temperatures up to 325°C[2]. These can be classified according to the layer coated in the inner wall of the column as:

(i) Wall–coated columns: where the stationary phase consists of a uniformly thin film which is usually a liquid of high–viscosity and low volatility, and the compounds are separated based on the partition process. The film thickness can vary from 0.1 to 0.25μM in thickness[2–3].

(ii) Porous layer columns: the stationary phase here is a porous layer. This is a solid which mainly operates according to the adsorption process[2–3].

In a GC system before the injection the column is being kept at a suitable temperature and the carrier gas is continuously passing through the column. Briefly, when the sample is introduced into the column inlet, each analyte is dynamically equilibrated and partitioned between stationary and the mobile phase. Nevertheless at any time the analyte distribution in the mobile phase is a function of its vapor pressure *i.e.* highly volatile (low boiling temperature) compounds will be partitioned to a greater extent toward the mobile phase and eluted first from the column while compounds with lower vapor pressures will be eluted later. So the efficiency of the separation will relay on the differences in the values of volatility from the compounds in the mixture.

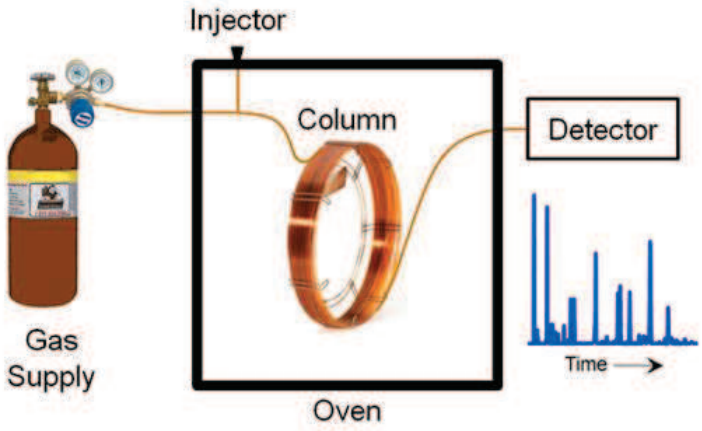


Figure 2.1. GC–system diagram taken from Agilent webpage http://www.agilent.com/labs/images

A schematic figure of the components of a GC instrument is shown in Figure 2.1. In a GC system, before the injection the column is being kept at a suitable temperature and the carrier gas is continuously passing through the column. Briefly, when the sample is introduced into the column inlet, each analyte is dynamically equilibrated and partitioned between the stationary and the mobile phase. At any time the analyte distribution in the mobile phase is a function of its vapour pressure i.e. highly volatile (low boiling temperature) compounds will be partitioned to a greater extent toward the mobile phase and eluted first from the column, while compounds with lower vapour pressures will be eluted later. Therefore, the efficiency of the separation will rely on the differences in the values of volatility from the compounds in the mixture.

The separation of a mixture can be optimized by increasing the number of changes in temperature (method with temperature gradient). Generally the isothermal analysis leads to peak broadening, enhancing with the time of the analysis. Additionally, as a secondary separation mechanism, the affinity

for the stationary phase by the analytes can promote their separation depending on physicochemical characteristics of both column and analytes, and temperature. These characteristics influence the separation of, for example, dipole–dipole energy, hydrogen bonds, and dispersion forces[2-4]. In brief, the separation mechanism in GC is based on the compounds' volatility and polarity.

Important factors that influence the chromatographic separation include column properties (length of the column, stationary phase, and internal diameter), oven temperature program, and carrier gas type and flow rate[4]. Fast GC without sacrificing chromatographic resolution can be achieved by using short– and narrow–bore capillary columns, fast column temperature programming and high carrier gas velocity[5].

One of the most critical aspects of GC is the injection process. Here, samples need to be introduced quickly, and only an injection volume between 1–5µL of solution. The injected sample is vaporized within the injection port which temperature is usually 50°C above the final column temperature. Here the non–volatile material accumulates on the glass insert and hence should be removed and cleaned at frequent intervals. Among the different types of injection i.e. split, splitless and temperature vaporized; split is the most used. The injection in split was the first method successfully applied for capillary GC analysis in which the sample size is made compatible with the low capacity of the column[3]. The two prerequisites of compounds to be measured by GC are volatility and thermal stability otherwise these requirements can be enhanced by chemical modification (derivative formation).

GC is a high resolution technique that can separate a large number of compounds in complex matrixes[6] and provides highly reproducible retention times (RT). GC has been extensively used for investigating volatile compounds in biological samples, where chemical derivatization procedures are now widely applied in the metabolomics[7]. The advantages and drawbacks of the derivatization process are presented in Table 2.1. Although the labeling process has some disadvantages it is a compulsory procedure for most of the biological samples where most of the compounds contained are not volatile.

Derivatization refers to the process of blocking the polar groups from molecules to improve their characteristics in terms of volatility, and thermal stability[8]. Generally, for labeling process two–step derivatization is often adopted. The protocol followed in the manuscript of this chapter is described in detail elsewhere[9]. Briefly, after metabolite extraction samples are dried to permit the action of reagents

following in the derivatization reaction. The first stage is called "*methoximation*", where the dried extract is dissolved with metoxiamine (CH_3ONH_2) in pyridine, stabilizing carbonyl moieties and reducing the number of derivatives from sugars. For derivatization *per se,* silylation is the most widely used protocol in which active hydrogen atoms are replaced with an alkylsilyl group, i.e. a trimethylsilyl group (TMS, –$SiMe_3$). Here, functional groups that present a volatility problem such as: hydroxyl (–OH), carboxyl (–COOH), amine (–NH_2), thiol (–SH), and phosphate (–PO_4^{3-}) are derivatized by silylation reagents. Importantly, while the (–OH) and (–COOH) groups react simultaneously and very fast, the hydrogen atoms of the (–NH_2) groups react slower and can yield multiple derivatives of some compounds[4, 9–10]. Among the silylation main pitfalls, the requirement for the sample to be dried and the limited stability of the derivatives when exposed to moisture in the air, are two important concerns[8].

Table 2.1. Characteristics of the derivatization procedure for metabolites: advantages and drawbacks[8, 11-13].

.Advantages	Weakness
▪ Allow volatility and stability of the metabolite to high temperature analysis	▪ It is a time–consuming procedure
▪ Improve chromatographic properties by reduction in adsorption and amplification of structural differences	▪ High variability and bias might be introduced
▪ Help for identification of chemical structure based in terms of numbers of reactive groups, such as hydroxyl and carbonyl functions	▪ High probability that some compounds of interest might be lost in the labelling process
▪ Introduce chemical selectivity to form diastereoisomers for chiral separations.	▪ Not all compounds can be derivitized
▪ For trained analysts this is an efficient, quantitative and reproducible methodology	
▪ Allows fragmentation of the analyte in the MS	

GC has had different types of detectors according to the broad applications that have been employed. These can be classified as universal or selective, depending on their detection capability. The universal detectors employed for GC are: (*i*) thermal–conductivity detector (TCD) which measures the difference in thermal conductivity between pure carrier gas and carrier gas plus a solute; (*ii*) flame–ionization detector (FID), which measures organic solutes by their oxidation in a hydrogen–rich flame between two electrodes; and (*iii*) mass spectrometry (MS). In addition, several selective detectors have been used in GC, among them electron–capture detector (ECD), nitrogen/phosphorus detector (NPD) and flame photometric detector (FPD) are some examples[2, 7].

2.2. Characteristics of the Mass Spectrometers (MS) used in Gas Chromatography (GC)

GC–MS was one of the first techniques to be coupled: its success during the sixties led to the idea of coupling MS with a separation technique[11]. This achievement was possible because both techniques worked with samples in a gas phase and required minimum sample volume. The problem was the pressure/vacuum difference between the two techniques, i.e., while GC works with a gas flow of 1–2mL/min, MS needs a low vacuum at entry to the ionization source (this is estimated in 10^{-3}Torr). This inconvenience was sorted out through the development of high–capacity vacuum generation systems called "*turbo–molecular pumps*" which are capable of maintaining the required vacuum for the coupling[3].

The MS instruments coupled to GC consist of: (*i*) sample introduction, (*ii*) ionization source, (*iii*) mass analyzer, (*iv*) ion detector and (*v*) data handling. Here, only the ionization source and mass analyser systems are going to be described next, the rest are fully described in Chapter 5.

(ii) Ionization source.

As the metabolites are already in the gas phase, the molecules could get charged by two processes:

1. *Electron impact ionization (EI).* By this process the molecules get charged and are fragmented in the source. EI achieves ionization by displacement of an electron from the

molecule where, as a result, the molecular ion ($M^{+\bullet}$) is produced (although not always observed because it undergoes decomposition to give charged fragmented ions[3]).

$$M + e^- \rightarrow M^{+\bullet} + 2e^-$$

EI is performed in a high–vacuum ion source (10^{-7} to 10^{-5}mbar, 200–250°C), where analytes in the gas state are bombarded with electrons from a heated filament and accelerated with a potential difference close to 70eV[4] (Figure 2.2). The ions produced by EI are in highly excited states and thus the energy is released by breaking the bonds between atoms, yielding fragments. The molecule will always present reproducible ruptures or fragmentation patterns, and even minimal instrument–to–instrument variability[3]. Thus, EI provides structural information and allows the creation of fragmentation spectral databases that, coupled with highly reproducible GC retention times (RT) (or retention indices, RI), allows for definitive metabolite identification. These mass spectral libraries are transferable between instruments regardless of manufacturer[12]. Additional advantages of this source are no ion suppression, so it's less prone to matrix effects and a highly suitable interface for GC. This type of ion source is the most widely employed in GC metabolomics for the non–target analysis approach[3].

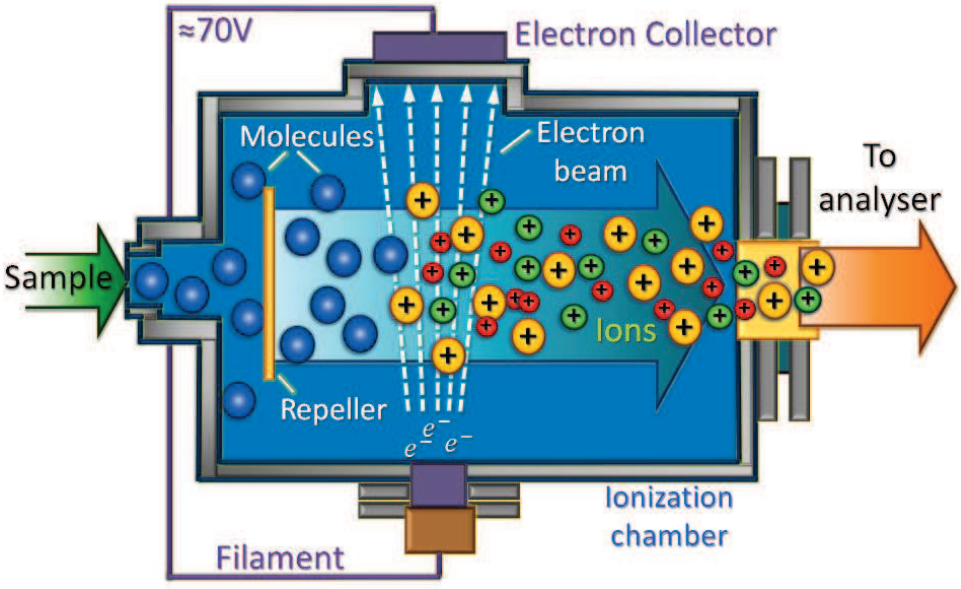


Figure 2.2. The ionization principle of electron impact (EI) adapted from http://www.chromedia.org/

2. *Chemical ionization (CI).* In this ion source, a high partial pressure of a reagent gas (e.g. methane, isobutane, or ammonia) previously ionized by electron impact is maintained in the ion source.

Then the ionization of the molecules is effected by reaction with reagent gas ions, leading to the generation of protonated molecule ions[3].

The reactive species after ion–molecule reactions in the gas are CH_5^+ from methane, and $C_4H_9^+$ from isobutane. This "soft" ionization produces more reagent gas ions capable of ionizing more molecules than electrons in the EI. In this ionization source more molecular ions than fragment ions are produced. The outcome is an improved signal for equal amount of analyte compared with EI ionization. CI is therefore used more for quantitative analysis and EI for qualitative analysis[3].

(iii) Mass Analysers.

After the ions are created, the next step is to transfer them to the mass analyser. The mass analyser is a part of the MS instrument that is responsible for separating ions according to their mass and charge (*m/z*). The MS mass analysers general characteristics are shown in Table 2.2. These vary in their capabilities along with their cost and the ease of use.

Table 2.2. Characteristics and performances of mass spectrometers commonly used in metabolomics[14].

	Quadrupole	Triple quadrupole	Ion Trap	Q–TOF	Orbitrap	Quadrupole–linear ion trap (Q–LIT)
▪ Mass accuracy	Low	Low	Good	Excellent	Excellent	Medium
▪ Resolving power	Low	Low	High	Very high	Excellent	Low
▪ Sensitivity (LOD)	Low	High	High	Medium	High	High
▪ Dynamic range	High	High	Medium	Medium	High	High
▪ MS/MS capabilities		✓	✓	✓	✓	✓
▪ Additional capabilities	–	Precursor, Neutral loss, SRM	Seq. MS/MS (MS^n)	MS/MS (MS^2)	Seq. MS/MS (MS^n)	MS^n, Precursor, Neutral loss.
▪ Identification	+	++	+++	+++	++++	+++
▪ Quantification	+	+++	++	++	++	++
▪ Throughput	+++	+++	+++	++	++	++

+, ++, +++, and ++++ respectively indicate *possible or moderate*, *good or high*, *excellent*, and *very high*. Seq., sequential.

A key aspect for the mass analysers is the ability of magnetic and/or electric fields to influence the movement of charged molecules in relation to their *m/z* under high vacuum. For example, two molecules that have identical trajectories and charge will behave differently under the influence of an electric and/or magnetic field if the molecules have different masses, i.e. heavier masses will be affected less than lighter ones. Therefore, by controlling the forces (magnetic and electric fields) the distinct molecules mass–to–charge ratio (*m/z*) can be obtained[15].

Specifically, the mass analysers available for GC instruments are: magnetic analyser, quadrupole (Q), ion trap (IT) and quadrupole–time of flight (QTOF)[2]. More specific characteristics of these are mentioned in Table 2.3. From all these the quadrupole was the analyser used in the GC–MS equipment for this chapter, and because of that, this is going to be described more extensively.

Table 2.3. Key Parameters of mass analyzers used in GC–MS instruments[2].

Parameter	Magnetic	Quadrupole	Ion trap	Time of flight
▪ Mass range (amu)	1–50,000	2–4,000	10–2,000	No limit
▪ Acquisition	Full–scan, SIM	Full–scan, SIM	Full–scan, SIM	Full scan
▪ Resolution	High 0.001 (*m/z* 1000)	Unit mass 1 (*m/z* 1000)	Unit mass 1 (*m/z* 1000)	High
▪ Mass accuracy	Accurate (1 ppm)	Nominal	Nominal	Accurate (5 ppm)
▪ Lower detection limit (g)	$10^{-15} – 10^{-14}$	$10^{-13} – 10^{-12}$	$10^{-13} – 10^{-12}$	–
▪ Linearity (order of magnitude)	3 – 4	4 – 5	3 – 4	–
▪ Scanning	Slow	Fast	Fast	Ultrafast

A quadrupole mass analyser (Q), as its name suggests consists of four precisely aligned metal hyperbolic or cylindrical parallel rods arranged symmetrically in a square array (Figure 2.3). Opposite pairs of electrodes are electrically connected and imposed direct current (DC) and oscillating radiofrequency (RF) electric fields to create conditions where ions of only a certain *m/z* (typically a 1*m/z* mass transmission window) will have a stable trajectory inside the device, and therefore pass through it. Inside the Q, the ion path starts as circular and gets transformed to complex spiral–like propagation inside the field. At a given value of DC and RF, only ions of a specified *m/z* have stable trajectories and are able to move through the Q mass filter and reach the detector, while all the other ions are lost by leaving the analyser prematurely or colliding with the rods and skeleton (Figure 2.3). Therefore, the Q mass analyser is a form of mass filter too. To obtain a mass spectrum, the Q field is changed by simultaneously scanning

DC and RF voltage amplitudes, allowing ions of each *m/z* to be passed sequentially through the Q and subsequently detected. Simultaneous increase in RF and DC amplitudes is necessary to transmit ions with larger *m/z*. Altering the optimal ratio will increase or decrease the *m/z* window, thus impacting the selectivity of detection and mass resolution [14-15].

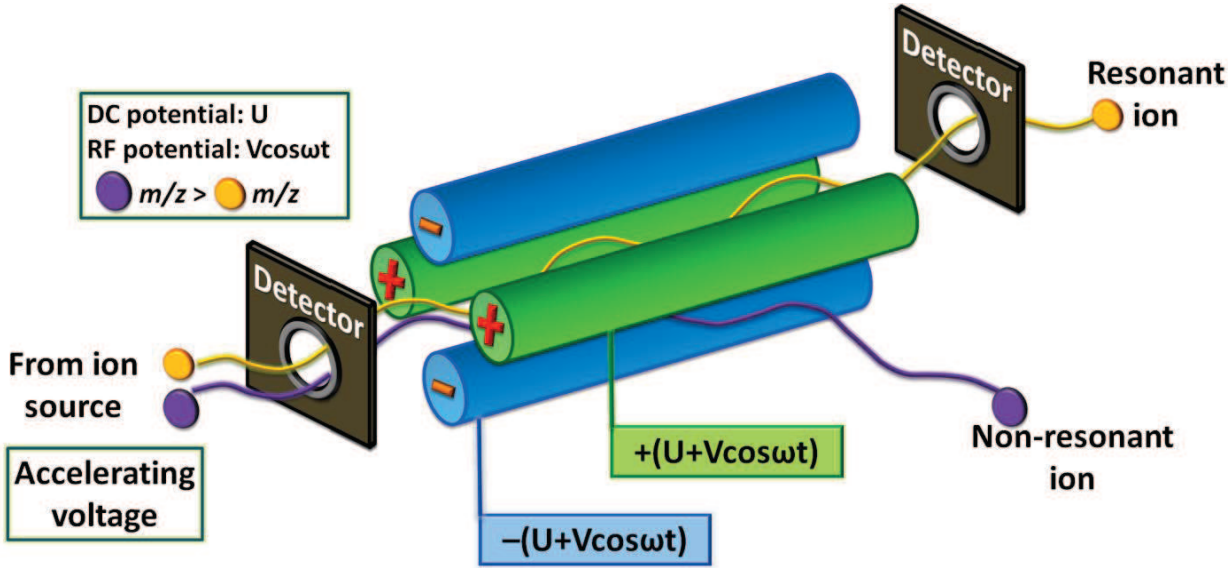


Figure 2.3. Schematic representation of a quadrupole (Q) mass analyser adapted from Tretyakova, N., 2012[14] and Rubakhin, S., 2010[15]. Here a Q field is created by applying a positive DC at a potential U with an imposed RF potential V *cos*ωt to the first electrode pair. The alternating pair of electrodes receives the negative DC potential (– U) and an RF potential which is out of phase by 180°.

Although the upper *m/z* range of the Q is not high, its use with IE ionization allows the detection of most of the charged fragments that reach the Q analyser. In addition, Q instruments are popular because of their high–linear dynamic range, rapid scanning capabilities, ease of control, size, and low cost[3, 15]. Nevertheless Q provides only nominal mass resolution[16]. Importantly in MS part, this should not be set to scan small mass fragments (less than *m/z* 50) because a high amount of noise may be encountered due to interferences like nitrogen (*m/z* of 28) in air[4]. They are widely used in GC–MS and a variety of hybrid instruments exist with multiple Q mass filters installed in the ion path, including systems

as complex as triple quadrupole (QqQ). Where the first Q selects a narrow *m/z* range, the second Q acts as a collision cell where the molecule can be fragmented, and the third Q scans and passes the resulting fragments towards the detector[15]. Therefore, the Q could work in different ways: it may act as a mass filter or mass selector; as a collision induced dissociation (CID) cell where ions are accelerated to a high kinetic energy resulting in collisions with a neutral gas, breaking molecular bonds and creating mass fragments[14]; and finally as a mass analyser.

Importantly, the use of mass analysers such as quadrupole–time–of–flight (QTOF)–MS coupled with GC allows for high–resolution and fast data acquisition, reducing the potential overlap of metabolite peaks and resolving co–eluting peaks. Additional Q–TOF acquires spectra for all of the ions at once, in contrast to scanning instruments such as Q[5].

2.3. Characteristics of GC–Q–MS in the non–target analysis

As the first hyphenated technique used in metabolomics, GC–MS was described as the gold standard in the metabolite profiling[7]. GC–MS is now a robust and very widely used technique where the combination of high resolution and reproducibility from GC and sensitive detection of MS have opened up extraordinary opportunities in biological research[4, 6, 11, 17]. Additionally, the use of electron impact (EI) ionization for GC–MS allowed the creation of spectral libraries which further facilitates the identification of biomarkers[4].

However as major shortcomings, GC–MS analysis requires compounds to be volatile and thermally stable, otherwise chemical derivatization must be performed. This means that for the analysis of polar compounds in biological samples, a relatively long time for sample preparation and instrumental analysis is expected. Furthermore, not all the biological compounds are amenable to derivatization[4]. This aspect has prompted several authors to describe GC–MS as a relatively low throughput technique for metabolomics applications, resulting in complex chromatograms due to the large number of metabolite peaks and multiple derivatization products[5, 11]. Clearly, then, the advantage of GC–MS lies in its high sensitivity of detection and identification for lot of volatile chemical compounds or non–volatile

compounds readily derivatized. These strengths and pitfalls of GC–MS for metabolic non–target analysis are summarized in Table 2.4.

Table 2.4. Relative strengths and weakness of GC–MS for metabolic non–targeted analysis[4].

Strengths	Drawbacks
▪ High chromatographic resolution, large peak capacity, high sensitivity, and reproducibility	▪ Limited to small compounds that are thermally stable, volatile, or can be made chemically volatile by derivatization e.g. sylilation
▪ Availability of EI reference mass spectra libraries and standard Kovats retention index (RI) for metabolite identification	▪ Extensive sample preparation (including derivatization usually >2h) and long chromatographic analysis
▪ Small sample size	▪ Identification capabilities until 4, 000Da

Therefore, the recent development and use of GC×GC/TOF–MS instruments combined with advances in chemometric data analysis have greatly enhanced resolution, improved sensitivity, increased peak capacity and the coverage of the metabolome, and the generation of unique ordered/structured chromatograms (although it has low reproducibility)[3, 7]. In addition, several works have shown the complementary use of GC–MS and other platforms such as NMR in the metabolomics field[11,18].

2.4. Insight into the application of GC–Q–MS as a complementary tool to LC–QTOF–MS

GC–Q–MS is one of the most frequently used tools for profiling primary metabolites. This type of instrument has proved to be mature and robust enough to run large sequences of samples for metabolomics studies with high chromatographic resolution and reproducible retention time (RT).

There are well established protocols for the analysis of most of the common biological samples such as urine and plasma by GC–MS[13, 19]. The importance of this technique is based on that not only

volatile and thermally stable compounds found in the biological matrix are going to be detected. Also small and polar metabolites which are amenable to be derivate will be detected; this includes metabolites such as amino acids, organic acids, carbohydrates, amines, amides, fatty acids, cholesterol among others.

Metabolic non–target analysis is extending beyond the most common biofluids (e.g. urine and plasma) to more specific but equally important ones like breast milk (BM). This biofluid is of paramount importance because it theoretically represents the only food that a newborn is receiving. Thus, BM is supposed to be ideal food that should cover baby nutrition, growth, and health[20–21]. Nevertheless, BM has not been characterized previously by any non–target analysis method. Therefore, one of the objectives of the next manuscript is the metabolic characterization of this biofluid by developing a dual strategy that could allow BM analysis in LC–QTOF–MS and GC–Q–MS, both from the same extracted sample aliquot obtained in one phase.

BM is a complex fluid where polar and non–polar compounds are together in a stable form, mostly as emulsified globules. In this sample most of the compounds contained are lipids such as triacylglycerol (TG)[22]. Typically normal GC has been used for the determination of free fatty acids (FFA) of the different lipid classes such as TG, lysophospholipids (LPL), cholesterol esters (CE) and so on. To reach the best organic extraction protocol that could include lipids for BM several methodologies where tested and analysed by LC–MS. Of importance, the methodology chosen had the advantage that the extraction system of methyl–*ter*–butyl ether/methanol (1:1) forms one phase, making the extraction procedure much simpler, faster and possible to split and measure in two analytical platforms at the same time. The extraction was characterized by both techniques LC–MS and GC–MS, this last through "*Fiehn GC–MS metabolomics retention time locks (RTL) method*".

The Fiehn GC–MS method is a commercially available methodology that allows an easy identification based on not only its EI spectral library created by Prof. Oliver Fiehn where the fragmentation pattern of the most common metabolites is present, but also based on their RT. The RTL tool permits the analyst to adjust the carrier gas pressure inside the GC instrument to achieve constant RT of the internal standard (IS). Results from this methodology at the specific extraction conditions allowed the detection of 60 metabolites in an extracted BM sample. Among them, amino acids, organic acids, FFA, hexose and pentose sugars, TCA intermediates, cholesterol and disaccharides were observed.

Importantly, disaccharide sugars were easily extracted and identified in the GC–MS profile. For example, lactose represents one of the highest peaks in the chromatogram. Nevertheless, during the experimental analysis it was observed that due to their high content in the sample the filament from the EI source got dirty quickly, and consequently the sensitivity of the analytical signal was decreased. So for method optimization, it was established to turn off the filament during the time that spans disaccharide molecules.

Due to the complexity of any non–target profile which typically contains from dozens to thousands of different metabolites, an extensive validation of the methodology seems to be compulsory in order to assess the overall performance of the method. In the case of GC–Q–MS, the comprehensiveness of the sample preparation; metabolite extraction and derivatization process, and the analytical performance are needed of validation. However, no guidelines on how to validate analytical non-target methods have been provided so far[23]. In this sense some authors have published some analytical strategies for validation. For example, Guy, P. in 2008, showed the within–day MS data reproducibility by using QC data to validate datasets from global profiles of urine samples by UHPLC–MS[24]. In addition, more authors have used the typical quantitative procedures in a usually limited and defined number of analytes; examples of this are: Tian, J., in 2009, showed the validation of some commercial compounds commonly found in bacterial cells for a non–target method characterizing bacterial metabolism by GC–Q–MS[25]; and Timischl, B. in 2008 showed the quantitation and validation of 20 metabolites in a CE–QTOF–MS method developed for a negatively charged metabolic profile, and applied the characterization of a mutant species from *Escherichia coli*[26]. Recently, Naz, S., in 2013 showed the validation of 22 metabolites spread among three analytical platforms (LC–QTOF–MS, GC–Q–MS and CE–TOF–MS) for the validation of metabolic fingerprinting of lung tissue[27].

Even after these examples it can be thought that there is not an effective way to validate completely a non–target analysis method. If the validation is done through the use of QCs samples this does not represent that the metabolites follow a linear response, accuracy or achieve the limit of quantitation. Otherwise in the validation based on standards, this is limited by the commercially available reagents and because only few of them can be tested (otherwise the work becomes infinite). This way does not represent all metabolites in the profile, but in order to consider a complete validation, parameters such as selectivity, calibration model (linearity and range), accuracy, precision (repeatability

and intermediate precision), and limit of detection (LOD) and quantification (LOQ) need to be tested. For this work 10 different metabolites were taken to be validated in the parameters mentioned above by GC–Q–MS.

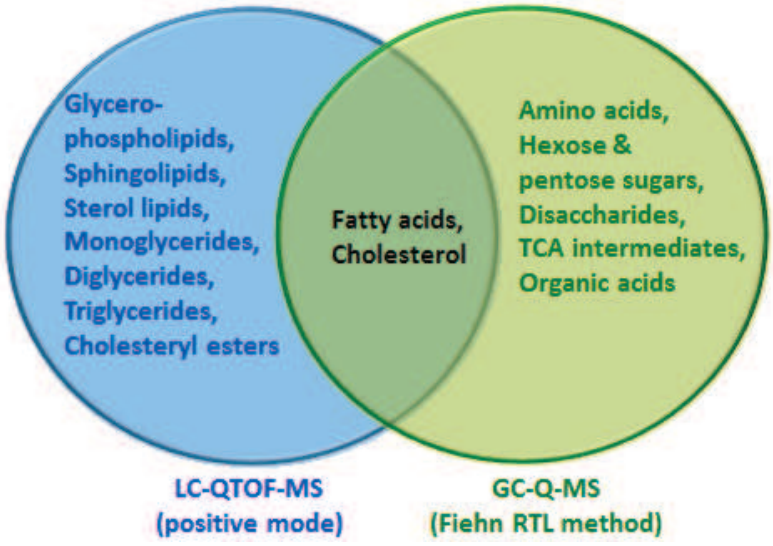


Figure 2.4. Venn diagram presenting data for biochemical classes detected in BM extraction by LC–MS and GC–MS.

For this chapter of the thesis the combination of two analytical platforms was again proved (as for Chapter 4, NMR + CE–UV). Data treatment for the combination of LC–MS and GC–MS is different than in the case of Chapter 4. Here both techniques are 3D data and have different type of MS and ion source, sample treatment and method of analysis which made it necessary for the data to be treated independently. Both give different and complementary information about the extraction system from BM. In this sense, a Venn diagram was done to compare the sort of biochemical compounds expected from each analytical platform (Figure 2.4). In this figure fatty acids and cholesterol were common metabolites from both platforms.

REFERENCES.

1. Snyder, L. R.; Kirkland, J. J.; Dolan, J. W., *Introduction to Modern Liquid Chromatography*. 3 ed.; John Wiley & Sons, Inc.: United States of America, 2010; p 960.

2. Cordero, C.; Liberto, E.; Sgorbini, B.; Rubiolo, P.; Bicchi, C., Gas Chromatography. In *Chemical Analysis of Food: Techniques and Applications.*, First ed.; Inc., E., Ed. Academic Press: USA, 2012; pp 311-373.

3. Honour, J. W., Gas chromatography-mass spectrometry. *Methods Mol Biol* **2006,** *324*, 53-74.

4. Pasikanti, K. K.; Ho, P. C.; Chan, E. C., Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** *871* (2), 202-11.

5. Han, J.; Datla, R.; Chan, S.; Borchers, C. H., Mass spectrometry-based technologies for high-throughput metabolomics. *Bioanalysis* **2009,** *1* (9), 1665-84.

6. Ramautar, R.; Demirci, A.; de Jong, G. J., Capillary electrophoresis in metabolomics. *Trends in Analytical Chemistry* **2006,** *25* (5), 455-466.

7. Issaq, H. J.; Abbott, E.; Veenstra, T. D., Utility of separation science in metabolomic studies. *J Sep Sci* **2008,** *31* (11), 1936-47.

8. Kaal, E.; Janssen, H. G., Extending the molecular application range of gas chromatography. *J Chromatogr A* **2008,** *1184* (1-2), 43-60.

9. Garcia, A.; Barbas, C., Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods Mol Biol* **2011,** *708*, 191-204.

10. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **2007,** *26* (1), 51-78.

11. Xiayan, L.; Legido-Quigley, C., Advances in separation science applied to metabonomics. *Electrophoresis* **2008,** *29* (18), 3724-36.

12. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Consortium, H. S. M., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2011,** *6* (7), 1060-83.

13. Chan, E. C.; Pasikanti, K. K.; Nicholson, J. K., Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc* **2011,** *6* (10), 1483-99.

14. Tretyakova, N.; Goggin, M.; Sangaraju, D.; Janis, G., Quantitation of DNA adducts by stable isotope dilution mass spectrometry. *Chem Res Toxicol* **2012,** *25* (10), 2007-35.

15. Rubakhin, S. S.; Sweedler, J. V., A mass spectrometry primer for mass spectrometry imaging. *Methods Mol Biol* **2010,** *656*, 21-49.

16. Lu, X.; Zhao, X.; Bai, C.; Zhao, C.; Lu, G.; Xu, G., LC-MS-based metabonomics analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** *866* (1-2), 64-76.

17. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot* **2005,** *56* (410), 219-43.

18. Teul, J.; Rupérez, F. J.; Garcia, A.; Vaysse, J.; Balayssac, S.; Gilard, V.; Malet-Martino, M.; Martin-Ventura, J. L.; Blanco-Colio, L. M.; Tuñón, J.; Egido, J.; Barbas, C., Improving metabolite knowledge in stable atherosclerosis patients by association and correlation of GC-MS and 1H NMR fingerprints. *J Proteome Res* **2009,** *8* (12), 5580-9.

19. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, C., Procedures for large-scale metabolic profiling of

serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2011,** *6* (7), 1060-83.

20. Miller, E. M.; Aiello, M. O.; Fujita, M.; Hinde, K.; Milligan, L.; Quinn, E. A., Field and laboratory methods in human milk research. *Am J Hum Biol* **2013,** *25* (1), 1-11.

21. Pons, S. M.; Bargalló, A. C.; Folgoso, C. C.; López Sabater, M. C., Triacylglycerol composition in colostrum, transitional and mature human milk. *Eur J Clin Nutr* **2000,** *54* (12), 878-82.

22. Jensen, R. G., Lipids in human milk. *Lipids* **1999,** *34* (12), 1243-71.

23. Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T., Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **2011,** *7* (3), 307-328.

24. Guy, P. A.; Tavazzi, I.; Bruce, S. J.; Ramadan, Z.; Kochhar, S., Global metabolic profiling analysis on human urine by UPLC-TOFMS: issues and method validation in nutritional metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** *871* (2), 253-60.

25. Tian, J.; Sang, P.; Gao, P.; Fu, R.; Yang, D.; Zhang, L.; Zhou, J.; Wu, S.; Lu, X.; Li, Y.; Xu, G., Optimization of a GC-MS metabolic fingerprint method and its application in characterizing engineered bacterial metabolic shift. *J Sep Sci* **2009,** *32* (13), 2281-8.

26. Timischl, B.; Dettmer, K.; Kaspar, H.; Thieme, M.; Oefner, P. J., Development of a quantitative, validated capillary electrophoresis-time of flight-mass spectrometry method with integrated high-confidence analyte identification for metabolomics. *Electrophoresis* **2008,** *29* (10), 2203-14.

27. Naz, S.; García, A.; Barbas, C., Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Anal Chem* **2013,** *85* (22), 10941-8.

Breast milk metabolome characterization in a single phase extraction, multiplatform analytical approach

FROM CEU: Alma Villaseñor[§], Antonia Garcia[§], Mariano Fernández-López[§], Coral Barbas [§]*.

[§]Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

FROM IMPERIAL: Isabel Garcia-Perez[1,2], Joram M. Posma[1], Andreas J Nicholas[3], Elaine Holmes[1]

[1]Computational and Systems Medicine Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK.

[2]Nutrition and Dietetic Research Group, Division of Endocrinology and Metabolism, Faculty of Medicine, Imperial College London, W12 0NN, UK.

[3]Section of Neonatal Medicine, Department of Medicine, Chelsea & Westminster Hospital, Imperial College London, London, SW10 9NH, UK.

AUTHOR INFORMATION

Corresponding Author

* Prof. Coral Barbas.

Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

Tel: +34 91 37 24 711

e-mail: cbarbas@ceu.es

Keywords: breast milk, biofluid, metabonomics, lipid, LC-QTOF-MS, GC-Q-MS

Abbreviations: BM, breast milk; LC-QTOF-MS, liquid chromatography-quadrupole-time of flight-mass spectrometry; GC-Q-MS, gas chromatography-quadrupole-mass spectrometry; GC, gas chromatography; LC, liquid chromatography; CE, capillary electrophoresis; MS, mass spectrometry; NMR, nuclear magnetic resonance spectroscopy; MTBE, methyl-*ter*-butyl ether; IS, internal standard; BSTFA, Bis(trimethylsilyl)trifluoroacetamide; TMCS, trimethylchlorosilane; MFE, Molecular Feature Extraction; MPP, Mass Profiler Profesional; AMDIS, Automated Mass Spectrometry Deconvolution and Identification System; EIC, Extracted Ion Chromatogram; PCA, principal component analysis; PLS-DA, partial least square discriminant analysis; FAME, fatty acid methyl esters; RI, retention index; QC, quality control; LOQ, limit of quantification; LOD, limit of detection; CE 18:2, cholesteryl linoleate; MG, monoglycerides; DG, diglycerides; TG, triglyceride; PG, glycerophosphoglycerols; PA, glycerophosphates; SNR, signal to noise ratio; LP1, lowest concentration of linearity; TCC, Total Compound Chromatogram; TIC, total compound chromatogram; TCA, tricarboxylic acid cycle; EI, electron impact source; RSD, relative standard deviation; PC1, first principal component; LPE, Lysophosphatidylethanolamine; LPC, lysophosphatidylcholine ; PE; phosphoethanolamine; PC, phosphocholine.

Total number of words:5 846.

ABSTRACT

Breast milk (BM) is a biofluid, which plays a fundamental role in early-life nutrition and directly impacts on growth and health. Global metabolic profiling is increasingly being utilized to characterize complex metabolic changes in biological samples. However, in order to achieve broad metabolite coverage, it is necessary to employ more than one spectroscopic platform, typically requiring multiple sample preparation protocols. In an effort to improve analytical efficiency and retain comprehensive coverage of the metabolome, a new extraction methodology was developed that successfully retains metabolites from BM in a single-phase using an optimized methyl-*tert*-butyl ether proportion. We conducted this single-phase extraction procedure on a representative pool of BM, and characterized the metabolic composition using LC-QTOF-MS in both and negative mode, and GC-Q-MS allowing characterization of polar and lipidic metabolites. To ensure that the extraction method was reproducible and fit-for-purpose, the analytical procedure was evaluated on both platforms using 18 metabolites selected covering a range of chromatographic retention times and biochemical classes. Having validated the method, the broader metabolic signature of BM composition was mapped as a metabolic reaction network highlighting interconnected biological pathways. Subsequently, the same protocol was applied to ascertain compositional differences between BM at week 1 (n=10) and 4 weeks (n=9) post-partum. Clear time-related differences were observed in both the GC-MS and LC-MS metabolite signatures reflecting differences in the relative concentrations of medium chain fatty acids (increased with time) and, sugar and cholesterol concentrations (decreased with time). This single-phase approach is more efficient in terms of time, cost, sample volume required and simplicity than the existing two phase methods, and will be suited to high-throughput metabolic profiling studies of BM.

INTRODUCTION

Breast milk (BM) is considered the best nutrition available for neonates; the World Health Organization recommends it as the sole source of nutrition in the first 6 months of life. BM has many functions other than supplying nutrition to neonates, containing many biologically active components, which have a fundamental role in infant growth and health[1-2]. As well as this, bioactive components found in BM are responsible for the metabolic development, immune and gut microbiome maturation of the newborn[3].

The major groups of nutrients present in BM are lipids, carbohydrates, proteins and many other micronutrients in smaller quantities[1,4]. BM is a highly variable fluid, whose composition also changes during lactation, and is highly influenced by maternal dietary intake[5-6]. Pre-clinical findings show that qualitative changes in dietary lipids during early life may contribute to long-term outcomes[7-8]. Previous studies[6] have emphasized lipid content is highly influenced by maternal nourishment. Jensen, et. al. described temporal differences in lipid composition, especially fatty acid content, over time postpartum, compositional variation from mothers with different ethnicities and diets, and between term and preterm milk[4].

Milk is a complex biological sample where fat is stabilized in the water phase by micelles due to the presence of several compounds acting as surfactants. After freezing, fat is agglutinated and the sample separates into two layers that are not stably mixed again after stirring or vortexing. Therefore, most published studies have adopted a two phase strategy and have analyzed the aqueous and organic phases separately. Given the complexity of BM, and its importance in human development, there is a need for development of global metabolic profiling methods that are capable of capturing as many nutrients and micronutrients present in the BM to give a comprehensive coverage of the metabolome with a view to establishing the influence of the chemical interaction between mother and infant via the breast milk.

Over recent years there has been extensive research into unraveling the metabolome of various biofluids[9]. The main aims of any global metabolic screening method is to achieve broad and rich coverage of as wide a range of endogenous and exogenous compounds in a single and robust profile, capable of representing the physiological condition of the subject studied[10].

Widespread analytical methodologies such as NMR spectroscopy and more recently GC, LC, and CE coupled to MS have proved valuable in the global assessment of metabolites in biofluids[11-13]. BM is an important fluid, however surprisingly there are two ^{1}H NMR global metabolic profiling studies reported in the literature[14-15]. In addition, most of the analytical research in BM to date has been targeted and focused primarily on the fatty acid content using liquid and gas chromatography[16-19]. Also other specific molecular classes such as the oligosaccharides[20] and sphingomyelins[21]; for example, Miller[20] et al have shown complex changes in structure and composition that is unique to mother-infant pairs and that the oligosaccharide composition directly effects the composition of the infant gut microbiome.

Apart from two studies using NMR[15] spectroscopy, to the best of our knowledge, no global MS-based assays have been used to study the global profile of BM. A robust screening method for BM capturing multiple classes of metabolites is necessary in order to carry out effective, high throughput screening of this biofluid. The aim of this study was to develop and optimize a methodology for analyzing BM that captures as many metabolites as possible using a single extraction phase in order to maximize the likelihood of identifying biomarkers associated with infant development and health.

We developed a single optimized organic phase extraction method for breast milk suitable for both GC-MS and LC-QTOF-MS and applied this method to the characterization of differences in BM composition at two distinct times post-partum.

MATERIALS AND METHODS

Reagents.

All standards, chemicals, reagents and solvents used in this study are fully described in the Supplementary Information (SI) and compounds used for the validation of both platforms are summarized in Table S-1.

Samples.

A pool of human breast milk was obtained mixing 100 µL aliquots from samples obtained from 52 women ranging from 1 to 76 days post-partum in order to characterize breast milk over the first four months of life. In addition, breast milk samples from mothers of healthy term-born infants were collected: in the 1st week (n=10) and in week 4 post-partum (n=9), and were analyzed in order to show the applicability of the method characterizing differences in BM composition over time. These two time points were chosen based that baby nutritional requirements are highly different while the newborn is growing. This application could be important to prepare special formulas for newborns more similar to the stages of breastfeeding. All samples were stored at -20°C until the day of analysis.

Sample extraction protocol.

Several extracting mixtures were compared in order to optimize the extraction protocol of as many metabolites as possible from the breast milk sample giving coverage of both lipids and polar metabolite classes.

Methanol-MTBE and ethanol-MTBE systems. 50 µL of BM was mixed with 350 µL of the solvents, methanol and MTBE in different ratios: methanol (0, 87, 175, 263, and 350 µL) and the converse volume of MTBE (350, 263, 175, 87, and 0 µL) were tested. The mixture then was vortexed for 1 min for protein precipitation and compound extraction. The same protocol was followed for the mixture ethanol: MTBE.

In all cases vitamin E acetate was used as internal standard (IS), at final concentration of 25 ppm.

GC-Q-MS sample treatment. Once the extraction protocol was optimized and established, analysis of the single phase extraction mixture by GC-Q-MS was performed. The analytical procedure for metabolic characterization was performed following a methodology previously described[22]. Briefly, an aliquot of 150 µL was transferred to a GC vial taken from the supernatant of the extracting system. The sample was evaporated to dryness in a Speedvac Concentrator (Thermo Fisher Scientific, Waltham, MA, USA). 10µL of O-methoxyamine

hydrochloride in pyridine (15 mg/mL) was added to each GC vial, and the resultant mixture was vigorously vortex-mixed for 5 minutes. Methoxymation was carried out in the dark, at room temperature for 16 h. For derivatization, the solution was vortex-mixed again for 5 min after 20 µL of BSTFA with 1% TMCS were added as catalyst. Samples were in an oven for 1 h at 70 °C for silylation process. Finally 100 µL of heptane containing 10 ppm of C18:0 methyl ester (IS) was added to each GC vial and vortex-mixed for 2 min before GC analysis (labeled sample).

A flowchart of the extraction protocol for the optimized method is presented in Figure 1.

LC-MS analysis method. Global profiling was developed using a LC-QTOF-MS instrument following methods described elsewhere[23]. The LC system consisted of a degasser, binary pump, and autosampler (1200 series, Agilent); 1 µL of extracted sample was injected into a reversed-phase column at 60°C (Agilent; Poroshell EC-C8, 15 cm × 2.1 mm, 2.7 µm) with a pre-column (Supelco; Ascentis Express C8, 0.5 cm× 2.1 mm, 2.7 µm). The gradient consisted of mobile phase A (10 mM ammonium formate in water) and mobile phase B (10 mM ammonium formate in methanol) pumped at 0.5 mL/min within a total run time of 60 min. 1 µl of extracted sample was injected to the column and eluted with a gradient starting at 75% B, increasing to 96% B in 23 min, then held until 45 min and increased to 100% B by 46 min, and held until 50 min. Starting conditions were returned by 51 min, and a 9 min re-equilibration time was included taking the total run time to 60 min. An online filter was added to the instrumentation. Data were collected in positive and negative ESI mode in separate runs on a QTOF analyzer (Agilent 6520) operated in full scan mode from 100 to 1200 m/z. The capillary voltage was 3500 V for positive and 4000 V for negative with a scan rate of 1.02 scans per second; the nebulizer gas flow rate was 10 L/min. During the experiment two reference masses were continuously detected allowing constant mass correction of data to obtain accurate mass in all injections [$(M_1\text{-}H)^+$= 121.0509 ($C_5H_4N_4$) and $(M_2\text{-}H)^+$= 922.0098 ($C_{18}H_{18}O_6N_3P_3F_{24}$)].

GC-MS analysis method. Labeled samples were analyzed in a GC instrument (Agilent 7890A) coupled to mass spectrometer with triple-Axic detector (Agilent 5975C). Two µL of sample volume was injected in split mode using an Agilent autosampler (7693). Split ratio was 1:5 to 1:10 with 3−10 mL/min helium split flow into a Restek 20782 deactivated glass wool split liner.

Separation was achieved using a 10 m J&W precolumn (Agilent Technologies®) integrated with a 122−5332G column: DB5-MS 30m length, 0.25 mm i.d. and 0.25 µm film consisted of 95% dimethyl/5% diphenyl polysiloxane (Agilent Technologies®). The helium carrier gas was used at a constant flow rate of 1 mL/min. The column was initially maintained at 60 °C for 1 min after injection, then temperature was increased at the rate of 10 °C/min to reach a final temperature of 325 °C, and cool down after analysis for 10 min. Temperatures of the injector, transfer line, filament source and quadruple were maintained at 250°C, 280°C, 230°C and 150°C, respectively. Quadrupole detector (5975 inert MSD, Agilent) was controlled by ChemStation software. The electron ionization source was operated at −70 eV. MS detection was achieved in scan mode over a mass range m/z 50−650 at a rate of 1 spectra/s. Filament of the EI source was turned off from minute 23 to 26. An *n*-alkane mixture from C8 to C28 dissolved in dichloromethane was run at the beginning of experimental worklist for calibration file and retention index determination. All these conditions were optimized previously[22].

For both techniques, the samples for the application were analyzed in one randomized run. QC samples were injected in the beginning and at the end of the sequence and in between each 5 samples to check system stability and performance of the analysis.

Data Treatment

LC-QTOF-MS. The resulting data file was cleaned of background noise and molecular entities were outgoing by the Molecular Feature Extraction (MFE) algorithm in the MassHunter Qualitative Analysis Software (Agilent®, version B.05.00). The resulting MFE file was a list of all molecular entities that accomplish the algorithm conditions, thus representing the full TOF mass spectral data in each sample. Primary data treatment (alignment and filtering) was performed in Mass Profiler Profesional (MPP) B.12.1 Agilent® software.

GC-Q-MS. Data was acquired with the Agilent MSD ChemStation Software. In GC-MS profiling, metabolite identification is a previous step for the statistical data treatment compared to LC-MS. Quality of chromatograms acquired by Total Ion Chromatogram (TIC) and internal standard peak were carefully examined. Automated Mass Spectrometry Deconvolution and Identification System (AMDIS) was used for automatic peak detection and deconvolution (see compound

identification). Compounds were identified by comparing their mass fragmentation patterns with those available in the NIST08 and Fiehn RTL mass spectral library.

For validation target analysis of compounds was performed: For LC-MS, the Extracted Ion Chromatogram (EIC) algorithm in MassHunter was used; this takes the *m/z* value of each compound and look for this into the chromatogram (±10ppm of mass error window). For GC-MS extraction and integration of a target fragment ion in MSD ChemStation Software was performed.

Data obtained from the analysis of 19 breast milk samples, obtained at two different time points, were processed using multivariate pattern recognition techniques using SIMCA P13.0.22.0 software (Umetrics, Sweden) such as principal component analysis (PCA), and partial least square discriminant analysis (PLS-DA)[24]. Unit variance scaling was applied to all data variables. The robustness of the models was evaluated based on R^2 (explained variance) and Q^2 (capability of prediction) values as well as 7-fold cross validation and class permutation validation. The regression coefficients from the PLS-DA models were divided by the jack-knife interval standard error (SE) to give an estimate of the t-statistic. Variables with a |t-statistic| ≥ 1.96 (z-score, corresponding to the 97.5 percentile) were considered significant[25].

Compound Identification

LC-QTOF-MS. Tentative identification of metabolites was done by searching mass accuracy against our own university online database CEU-mass mediator (http://ceumass.eps.uspceu.es/mediator/), which uses KEGG, METLIN and LipidMAPS databases with an error mass set to 10 ppm. Potential hits were compared to the experimental isotopic pattern distribution on MassHunter.

GC-Q-MS. A standard solution (Grain FAME mix®) was analyzed along the fingerprint worklist from which ten different fatty acid methyl esters (FAME) were chosen covering most of the retention time from the chromatogram. The FAME retention times (Rts) were used to create a calibration data file for further adjustment of samples Rts implying the accuracy improvement

in the determination of the Kovats retention index (RI). A well-established RI value is contained in Fiehn RTL library and this is compared to the experimental RI value in order to assign a match score between the experimental and the theoretical spectra[26]. Peak detection and deconvolution was performed by retention index comparison of spectra with the Fiehn RTL library in AMDIS software v.2.69[27]. Target compounds were assessed by comparing their mass fragmentation patterns with those available in the NIST mass spectral library and Fiehn RTL library. A private library was created with the targets obtained and confirmed from retention index search correcting retention time to the experimental observed. Metabolites that were not included in Fiehn RTL library such as cholesterol and arachidonic acid were added from the NIST library. Deconvolution was optimized by a second analysis performed by the use of retention time algorithm in AMDIS. Compounds present in at least 70% of all samples (QC samples + study samples) were kept for statistical analysis.

Validation methodology.

The method was validated to assess selectivity, linearity, accuracy, limit of detection (LOD), limit of quantification (LOQ), instrumental precision and method precision (last two with standards and samples). Compounds were selected by picking metabolites along the profile and which could cover different biochemical classes. For LC-QTOF-MS (8 metabolites): carnitine, alpha-linoleic acid, palmitic shpingosine, stearic acid, vitamin E, cholesterol, cholesteryl linoleate (CE 18:2) and triglyceride C18:2 (TG 18:2; TG(18:2/18:2/18:2)) were chosen whereas for GC-Q-MS (10 metabolites): citrate, glucose, glutamate, glycerol, lactate, myoinositol, proline, serine, stearic acid and urea were selected. Only stearic acid was the included as a standard for both techniques.

Linearity of the method was assessed by assaying five different levels of concentrations of standard solutions by triplicate covering the ranges of 25 to 200% of mean values for each metabolite .Recovery was estimated by comparing in triplicate, the values of spiked samples prepared in a linear range. Within-day instrumental and method precision were evaluated by consecutive injections (n= 10) of a homogeneous standard solution and pool of BM

respectively, prepared in the midrange of the calibration curve. Inter-day precision was tested on a different day, with freshly prepared standard solutions and new-no defrost BM pool. LOQ was estimated calculating signal to noise ratio (SNR) in the lowest concentration of linearity (LP1).

RESULTS AND DISCUSSION.

Sample treatment and extraction protocol.

After testing several extraction mixtures with different solvents, proportions and conditions, extraction protocols used methanol/ethanol: MTBE performed better than methanol:H2O:MTBE system (36: 21: 43%) in terms of giving a single phase containing a wide range of chemical classes. Here we present the optimized method.

Extraction procedures proposed and tested here were chosen based on minimum manipulation of the sample and broad and reproducible profile of metabolites in a single phase. MTBE was chosen primarily because it is the second solvent used after Folch method. Previous publications[23, 28] have shown that employment of MTBE avoids drying and re-suspension steps reducing manipulation. The difficulty of analyzing BM is the high content of lipids intrinsically packed in the sample which normally results in the formation of two phases therefore sample treatment cannot be avoided. The extraction procedure was selected based on comparing the results obtained from: (1) visual analysis of Total Compound Chromatogram (TCC), (2) IS recovery and (3) total number of entities (Figure 2).

Methanol:MTBE and ethanol:MTBE systems. All ethanol/methanol: MTBE mixtures showed one unique phase instead of the two phases commonly observed with solvent extraction. Methanol and ethanol work as co-solvents allowing miscibility of MTBE with the water in the milk. These mixtures had the advantage of forming one phase, which not only made the procedure simpler but also reduces analytical variation and avoids the partition phenomenon of compounds when two phases are formed. The chromatograms obtained in the methanol-MTBE serial showed that with the increase of MTBE proportion, the polarity of the mix solvent is diminished and therefore the amount of non-polar metabolites and intensity is higher (Figure 2A). From

methanol: MTBE proportion (50:50) to MTBE 100% differences were observed mainly at the beginning of the chromatogram.

The selection of the extraction procedure was based on the best recovery of compounds observed through the chromatograph, total signal and SI recovery. EIC of vitamin E acetate was used as the IS [m/z= 473.3989], this is the synthetic compound of endogenous vitamin and had a partition coefficient (log P) of 10.69 (http://www.chemspider.com/). Therefore, vitamin E acetate denotes the hydrophobicity and partition power in the media, and allowed ascertaining the recovery of a known compound with a medium polarity. As shown in Figure 2B-C, the IS and total signal (TS) were used to compare the extraction methods in order to select the best solvent proportions. Figure 2B-D showed that highest recovery of IS and TS compared to others systems was achieved with methanol: MTBE (50:50).

Intentionally, an additional extraction system was tested, this was the mixture of methanol: H_2O: MTBE (125: 75: 150 µL)[23] to force the formation of two phases in the resultant systems, allowing comparison of the one- and two-phase extraction procedures. Importantly, the chromatogram from the organic phase of methanol: H_2O: MTBE (Figure 2D) system showed lower intensity and less amount of compounds compared to one phase systems. Apparently the formation of the two phases in the system provoked that compounds the effect of partitioning in the two phases of the system.

Finally, the extraction with methanol: MTBE (50:50) was selected for global signature profile. From this point the characterization of the profile by the LC–QTOF–MS and GC–Q–MS was performed accordingly to the optimized method described in the flowchart in Figure 1.

Profile characterization.

LC-QTOF-MS. Metabolite characterization in the chromatogram was performed using the masses with a RSD less than 20% coming from 10 independent replicates of a BM pool. In addition, features with abundance less than 10^5 and 10^4 for positive and negative mode respectively were discarded reducing the dataset dimensions for further database identification (682 out of 1283 for positive and 204 out of 938 for negative mode). Hits from data bases were

clustered basically within a biochemical class and were depicted in a Total Compound Chromatogram (TCC) in Figure 3. These are summarized in Table S-2. The retention time window for the major classes of lipids shown in Figure 3A, matches the order of elution described with previous lipidomics studies[23, 29]. Among the group of compounds previously reported[4,6,21,30-32], fatty acids, glycerophospholipids, sphingolipids, sterol lipids, monoglycerides (MG), diglycerides (DG), triglycerides (TG) and cholesteryl esters were observed. As expected, fewer number of compounds were detected in negative mode. Primary fatty acids, and phospholipids were detected in both ionization modes, and carbohydrates (mono and disaccharides) in the dead-volume in negative mode only (Figure 3B). Interestingly Table S-2 shows some additional classes of compounds that are not described in literature for BM, but were tentatively identified such as fatty esters, fatty aldehydes, ceramides, glycerophosphoglycerols (PG) and glycerophosphates (PA).

Extracted ion chromatograms from the most common fatty acids; myristic, palmitic, stearic, oleic, linoleic, α-linoleic and arachidonic acid[7] in BM reported in literature were presented in Figure S-1. Here, the comparison of these 7 fatty acids showed different retention times in the chromatographic system. Fatty acids with longer carbon chain eluted later, while molecules with higher number of double bonds were associated with shorter retention time.

GC-Q-MS. The breast milk extraction was characterized using the standard "Fiehn metabolomics retention time lock (RTL)" method. The total compound chromatogram (TIC) with the metabolite assignments for 60 compounds is presented in Figure 5 and Table S-3. Metabolite classes detected by GC–Q–MS included; amino acids, organic acids, fatty acids, hexose and pentose sugars, tricarboxylic (TCA) intermediates, cholesterol and disaccharides. Disaccharides were easily extracted and identified in the GC-MS profile. Here, for example, lactose represents one of the highest peaks in the chromatogram. Nevertheless, during the experimental analysis it was observed that due to their high content in the sample the filament from the EI source got dirty quickly and consequently the sensitivity of the analytical signal was decreased. So for method optimization, it was established to turn off the filament during the time that spans disaccharides in the chromatogram from 23 to 26 minutes. Most of the BM

compounds identified were derived from maternal endogenous metabolism. However, heptadecanoic acid is synthesized by intestinal gut bacteria[33] from maternal diet and illustrates the early relationship between gut bacteria and human metabolism. Other literature has focused on oligosaccharides; such as fucose, and has shown that the primary purpose of BM oligosaccharides is to provide a nutritional source for the infant gut microbiota[20].

Method validation.

Although there are no guidelines about method validation for fingerprinting in the metabolomics field, a traditional validation method for bioanalysis was performed on 10 compounds for GC-MS and 8 metabolites for LC-MS. The metabolites were chosen to cover a range of biochemical classes, polarities, functional groups, molecular weight and retention time spanning the chromatograms in both techniques. A summary of the validation parameters is provided in Tables S-4 and S-5 for LC–QTOF–MS and GC–Q–MS respectively. For linearity, standards were fitted to the linear model ($r > 0.99$) for all selected metabolites, and no bias was found for most of them excluding; palmitic sphingosine and cholesterol for LC-MS, whereas a linear fit for urea and stearic acid was not obtained for GC-MS. No practical consequences were seen in the recovery yield for any metabolite despite the bias in the four metabolites mentioned (recoveries ranged from 90.7 to 106.1%). Specifically for CE 18:2 and carnitine measured by LC-MS the recovery yield was 51.5 and 63.8% respectively, thus both denoted the matrix effect owing to the other compounds in the sample at these particular conditions. However, although recovery was did not meet the validation criteria for these two compounds, they did meet the rest of rest validations parameters.

For LC-MS standards, instrumental precision (n= 10) ranged RSD from 2.8 to 7.0% last one from TG 18:2, which shows the importance of temperature in the sample bench and stability of the LC system which we suggest should be around 15 degrees to avoid lipid precipitation. Intra-assay precision for standards ranged RSD from 1.4 to 5.0% (n= 7) and inter-assay precision from RSD 1.9 to 5.0% (n= 14). For samples intra-assay precision: 7 samples were prepared from the same pool, treated from the beginning and run in the same assay daily for 2 days. RSDs ranged

from 1.4 to 5.8% on the first day. RSD using both days ranged from 3.9 to 8.2% (n = 14).Finally, the theoretical LOQ for LC-MS compounds ranged from 15 ppb for palmitic shpingosine to 389ppb for stearic acid.

In the case of the GC-MS standards, instrumental precision for the 10 chemical standards ranged from 2.8 to 7.7%. For standards intra-assay precision ranged from 2.8 to 7.9% (n= 7) and inter-assay precision from 3.6 to 8.6% (n= 14). For samples intra-assay precision RSDs values ranged from 3.0 to 6.8% and from 3.9 to 10.8% in different days (n= 14). Theoretical LOQ ranged from 0.4 ppm for proline to 15.439 ppm for glycerol. There was good agreement between the BM concentration of stearic acid measured by GC-MS (20.02 ± 1.73ppm) and LC-MS (21.00 ± 1.22ppm) showing reproducibility across analytical platforms.

Network analysis of breast milk composition

Having validated the single-phase extraction method for both GC-MS and LC-MS analysis of BM, the metabolic reaction network of the metabolites identified in BM was explored using the MetaboNetworks software[34]. The metabolic reaction network highlights the interconnectivity of different domains of metabolism by considering the shortest connecting paths found in KEGG that link the identified metabolites, it shows the composite metabolic signature of BM (Figure5). Metabolites shown in blue were identified using LC-MS, metabolites identified using GC-MS in red and metabolites identified by both LC-MS and GC-MS in magenta. The white nodes represent the intermediary metabolites in pathways connecting the identified compounds, these may play an important role in the internal intermediary metabolism hence they are not the final products detected by GC- and/or LC-MS analysis. In general, specific metabolite classes were detected mainly by one of the two analytical platforms employed and often multiple metabolites were detected per biological class. For example, branched-chain amino acids, metabolites related to glycolysis and amino sugars metabolism, aromatic compounds as well as TCA cycle intermediates and TCA anaplerotic metabolites were identified with GC-MS. On the other hand, Figure 5 highlights that metabolites related to lipid metabolism were identified by LC-MS. Cholesterol and related sterols (related to bile acid metabolism) were

also found mainly with LC-MS. This shows that each technique identifies mostly metabolites associated with specific pathways and highlights the necessity of analyzing BM samples using different analytical techniques in order to achieve maximal coverage of the BM metabolome. However, for certain compound classes, such as short chain fatty acids, both techniques were capable of detection.

Global profiling analysis on breast milk samples composition over time

The validated single-phase extraction method with LC- and GC-MS detection was applied to characterizing compositional differences in BM obtained 1 week and 4 weeks post-partum. Clear metabolic differences between the two time points based on spectral data obtained from LC-MS analysis were easily visualized using Principal Components Analysis (PCA). The PCA scores plot of the first three components corresponding to the LC-MS data demonstrated an evolving trend in the composition of the milk along the first principal component (PC1) in relation to its time of collection post-partum (Figure S2-A). However, the corresponding PCA scores plot for the GC-MS data did not show any sample clustering in any of the first three components in relation to its time of collection (Figure S2-B).

A supervised approach based on partial least squares discriminant analysis (PLS-DA) was subsequently applied to the data obtained by LC-MS (Figure S2-C) and GC-MS (Figure S2-D) using the two sample collection time points as classifiers. Here, both the GC-MS and LC-MS PLS-DA models showed clustering of samples according to class; with blue dots representing BM samples collected on week 1 post-partum and red dots representing samples obtained 4 weeks post-partum. This post-partum evolution in the composition of the milk is concordant with published literature[35-36]. To identify the metabolites which changed in relation to time post-partum, causing the separation between groups, loading plots of both PLS-DA models were explored. These models identified a number of different metabolites changing over lactation time. The data from LC-MS identified several metabolites that increased in concentration in the samples taken more than 26 days post-partum such as linoleic acid (18:2), palmitoleic acid (16:1), oleic acid (18:1), LPE (18:2), hydroxyadipic acid, and MGs, DGs and TGs. Metabolites

decreasing in concentration over this period were: lyso-and phospholipids, α-tocopherol, cholesterol and CE 18:2.

The loading plots of the PLS-DA model corresponding to the GC-MS data show oleic acid (18:1), palmitoleic acid (16:1), linoleic acid (18:2) and gluconic acid to be increased in concentration in samples collected over week 4 compared to the first 7 days post-partum. Metabolites decreasing in this period were: fucose, furanose (fructose), D-glucosaminic acid and cholesterol.

The ability to establish differences in the relative abundances of metabolites present in samples, in agreement with previous literature gives further weight to the validity of the method developed and outlined in this paper. For example, previous research has similarly identified a decrease in the abundance of cholesterol from BM samples taken as lactation proceeded[35], as identified by GC-MS and LC-MS in this study. Here, the total amount of cholesterol and cholesterol ester declined studied over a period of 2-84 days postpartum. cholesteryl esters were observed to decrease from approximately 5mg/dl to 1mg/dl.

Likewise, increases in the quantity of certain fatty acids have been described in BM samples collected at different times during lactation[36], in line with our findings. One such study examined fatty acid abundances in breast milk collected from day two to day 84 post-partum. Oleic acid 18:1, was found increased from 36.08± 1.03 on day two to 38.67 ± 0.84 on day 84 post-partum[36]. Similar findings were reported for the linoleic acid 18:2, with the percentage weight increasing from 13.16 ± 0.80 to 17.46 ± 0.67 over the entire period studied[36]. These literature results are all reflected by our findings from a combination of GC-MS and LC-MS, see Tables S-6 and S-7.

Concluding remarks

We have shown for the first time how BM samples can be analyzed using a single phase extraction followed by the global profiling analysis of the same phase using GC-Q-MS and LC-QTOF-MS, which enables simultaneous characterization of both medium polar compounds and lipids simultaneously. Our approach is an improvement over existing methods for the analysis of BM that typically use a two phase extraction as it is more efficient in terms of time, cost and

simplicity, and serves two different analytical platforms. We have applied this method to the characterization of the human BM metabolome and have used pathway mapping tools to identify representation of multiple interconnected pathways. These pathways include branched chain amino acids, aromatic amino acids, TCA cycle intermediates and anapleurotic metabolites, short chain fatty acid metabolism and lipid metabolism. Unsurprisingly, the metabolic reaction network shows that metabolites from a specific pathway are mostly identified using a single technique. However, in the case of metabolic pathways that are co-regulated by gut bacteria, both GC-MS and LC-MS indicate connectivity. In future, analyzing the metabolic profile of breast milk using different analytical techniques will be needed to uncover the full breast milk metabolome. Finally applicability of these proposed analytical methodologies was corroborated after studying changes in breast milk composition over time.

Acknowledgments

AV acknowledges EADS-CASA for her fellowship. Authors gratefully acknowledge the financial support from Ministry of Science and Innovation MICINN CTQ2011-23562. IGP is a NIHR postgraduate research fellowship.

Conflict of interest statement

The authors declare no competing financial interest.

Figures

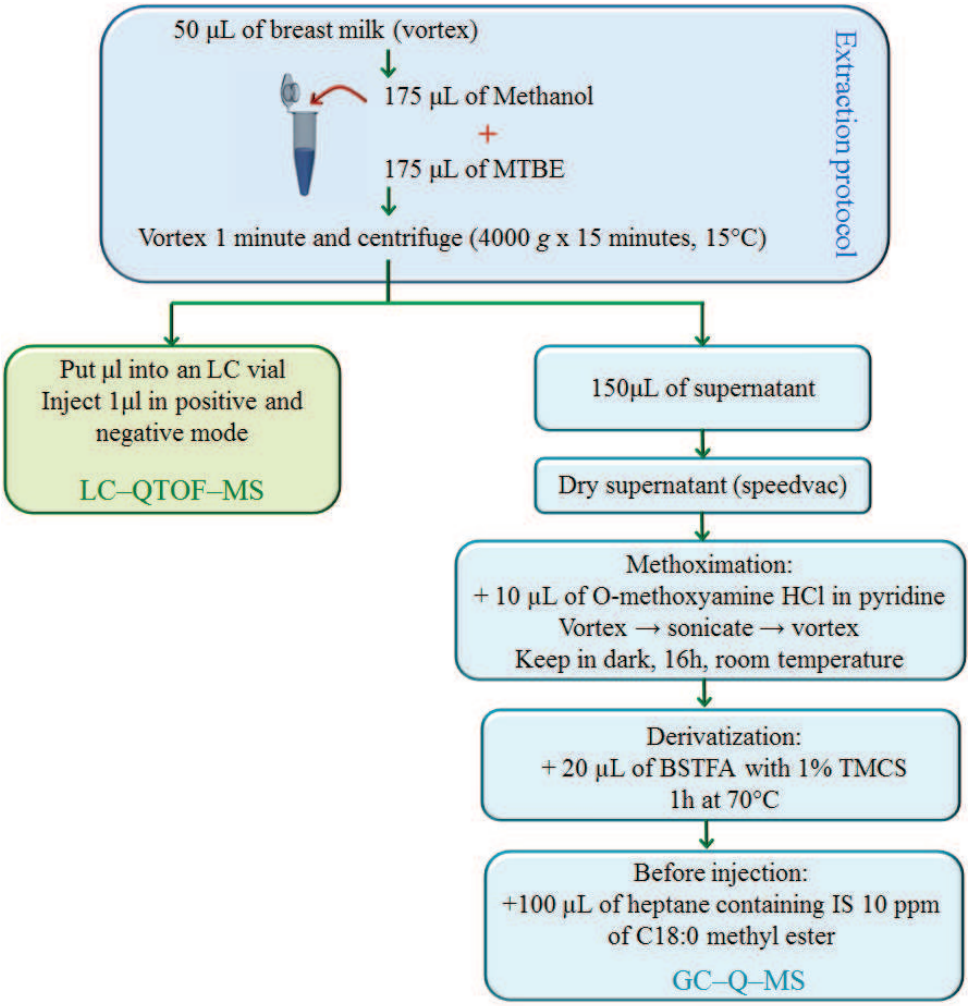
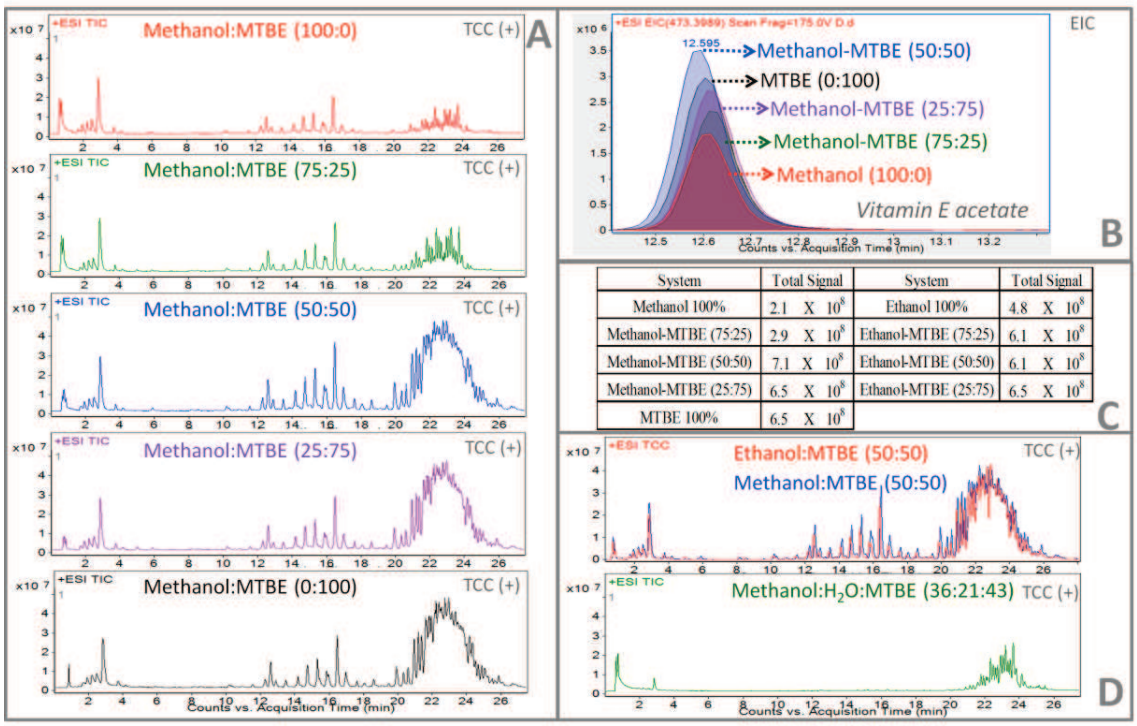


Figure 1. Optimised breast milk extraction protocol flowchart for LC–QTOF–MS and GC–MS.



System	Total Signal	System	Total Signal
Methanol 100%	2.1 X 10^8	Ethanol 100%	4.8 X 10^8
Methanol-MTBE (75:25)	2.9 X 10^8	Ethanol-MTBE (75:25)	6.1 X 10^8
Methanol-MTBE (50:50)	7.1 X 10^8	Ethanol-MTBE (50:50)	6.1 X 10^8
Methanol-MTBE (25:75)	6.5 X 10^8	Ethanol-MTBE (25:75)	6.5 X 10^8
MTBE 100%	6.5 X 10^8		

Figure 2. Graphical comparison of the different extraction mixtures tested using positive mode ionization. (A) LC-QTOF-MS Total Compound Chromatograms (TCC) from methanol (100, 75, 50, 25, 0%) - MTBE (0, 25, 50, 75, 100%) proportions; (B) Extracted Ion Chromatogram (EIC) from Vitamin E acetate (m/z= 473.3989) for the extraction mixtures aforementioned in (A); (C) Total Signal (TS) of all methanol / ethanol-MTBE proportions; (D) TCC comparison of methanol / ethanol – MTBE (50:50) and the TCC of the extraction mixture of methanol:H_2O-MTBE (36:21:43).

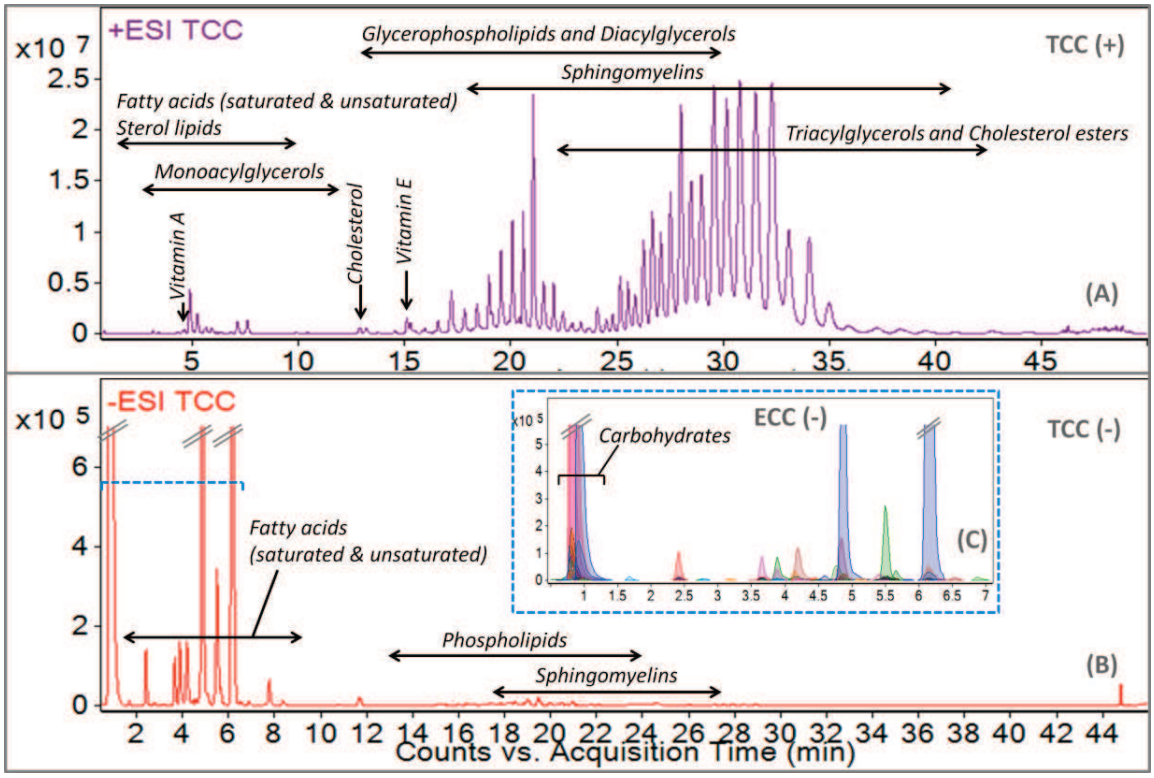


Figure 3. (A) Total Compound Chromatogram in positive mode (TCC +) and (B) TCC (-) in negative mode, acquired by LC-QTOF-MS. Metabolite identification in BM profile by classes of lipid. (C) Extracted Compound Chromatogram (ECC) for negative mode showing carbohydrates. Extraction system [methanol: MTBE (50:50)].

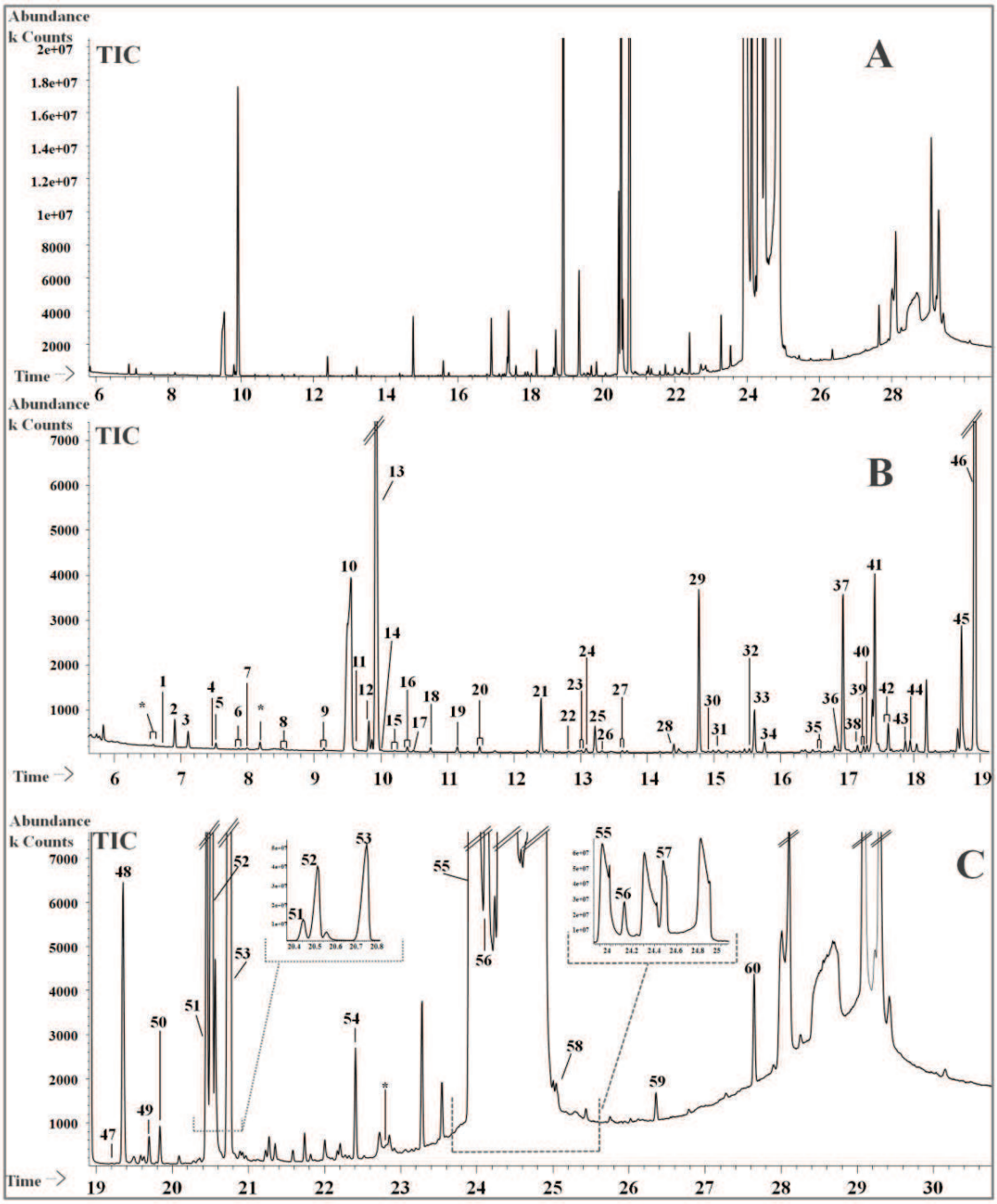


Figure 4. (A) Total Ion Chromatogram of breast milk profiling obtained by GC–Q–MS. (B) Zoom profile from 6 to 19 minute. (C) Zoom profile from 19 min to 31 min. Lipid extraction [methanol: MTBE (50:50) proportion], for more explicit conditions see the text. Numbers correspond to final identified compounds described in Table S-3.

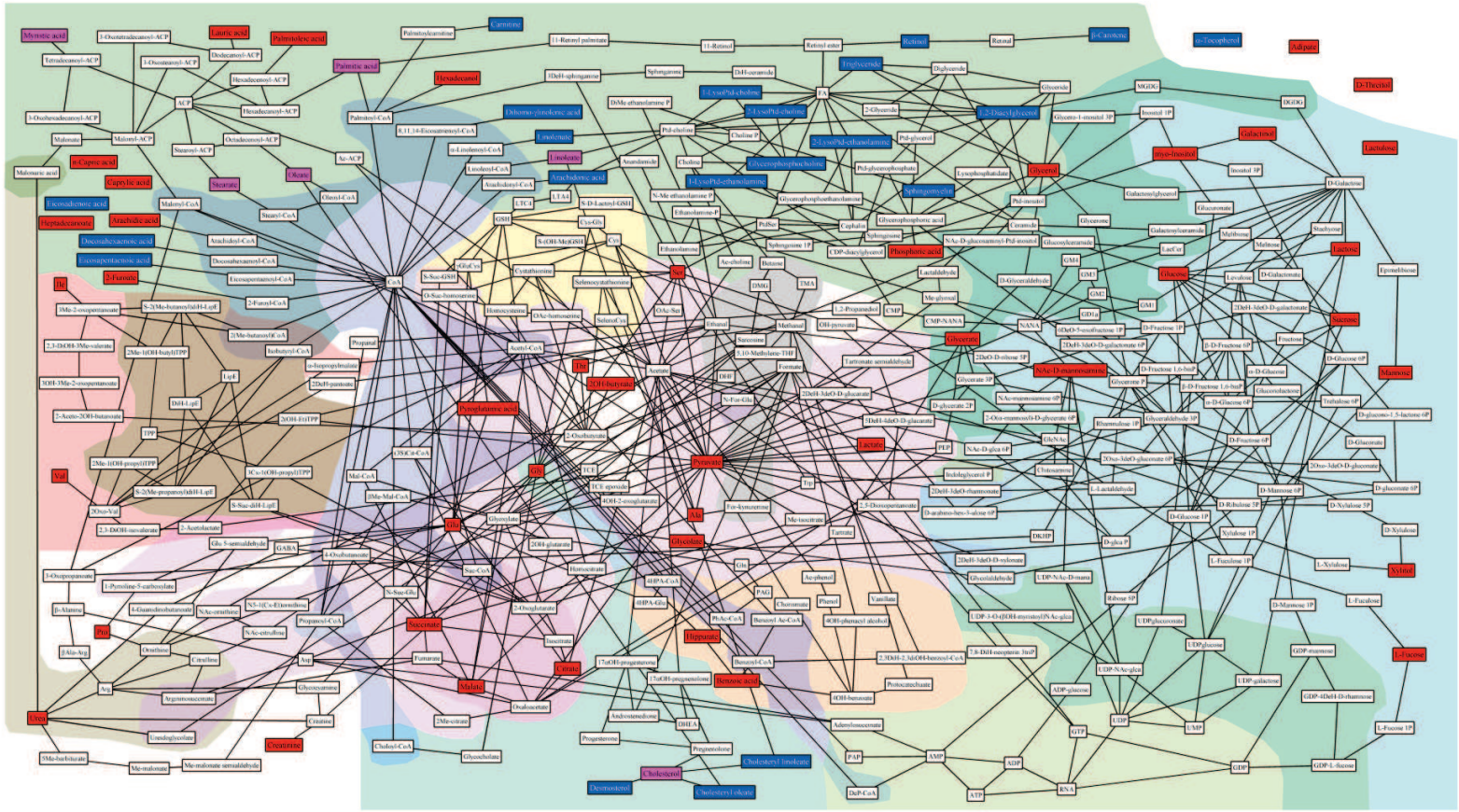


Figure 5. Metabolic reaction network of metabolites found in breast milk, created using the MetaboNetworks software (Posma et al, 2014). The network shows links between metabolites if the reaction entry in KEGG shows that it is a main reactant pair and the reaction is either mediated by an enzyme linked to human genes, enzyme linked to genes from the most abundant endosymbionts (the phyla firmicutes, bacteroidetes, *alpha*-proteobacteria, *beta*-proteobacteria, *gamma*-proteobacteria, *delta*-proteobacteria and

actinobaqacteria) or it is part of a spontaneous process. Metabolites shown in blue were identified using UPLC-MS, in red metabolites identified using GC-MS, in magenta identified by both LC-MS and GC-MS and in white are all metabolites linking these metabolites. The background shading illustrates different types of metabolism based on the closest affinity with some overlap between groups: tricarboxylic acid (TCA) cycle ( ); TCA anaplerotic metabolism ( ); coenzyme A metabolism ( ); lipid and fatty acid related metabolism ( ); branch-chain amino acid metabolism ( ); bile acid metabolism ( ); aromatic compounds ( ); sulphur metabolism ( ); purine and pyrimidine metabolism ( ); 1-carbon and related metabolism ( ); glycolysis and amino sugar metabolism ( ); and urea cycle metabolites ( ). An expanded version is provided in the online supplement.

REFERENCES

1. Miller, E. M.; Aiello, M. O.; Fujita, M.; Hinde, K.; Milligan, L.; Quinn, E. A., Field and laboratory methods in human milk research. *Am J Hum Biol* **2013,** *25* (1), 1-11.
2. Pons, S. M.; Bargalló, A. C.; Folgoso, C. C.; López Sabater, M. C., Triacylglycerol composition in colostrum, transitional and mature human milk. *Eur J Clin Nutr* **2000,** *54* (12), 878-82.
3. Oozeer, R.; van Limpt, K.; Ludwig, T.; Ben Amor, K.; Martin, R.; Wind, R. D.; Boehm, G.; Knol, J., Intestinal microbiology in early life: specific prebiotics can have similar functionalities as human-milk oligosaccharides. *Am J Clin Nutr* **2013,** *98* (2), 561S-71S.
4. Jensen, R. G., Lipids in human milk. *Lipids* **1999,** *34* (12), 1243-71.
5. Rasmussen, K. M., The influence of maternal nutrition on lactation. *Annu Rev Nutr* **1992,** *12*, 103-17.
6. Koletzko, B.; Rodriguez-Palmero, M.; Demmelmair, H.; Fidler, N.; Jensen, R.; Sauerwald, T., Physiological aspects of human milk lipids. *Early Hum Dev* **2001,** *65 Suppl*, S3-S18.
7. Abrahamse, E.; Minekus, M.; van Aken, G. A.; van de Heijning, B.; Knol, J.; Bartke, N.; Oozeer, R.; van der Beek, E. M.; Ludwig, T., Development of the Digestive System-Experimental Challenges and Approaches of Infant Lipid Digestion. *Food Dig* **2012,** *3* (1-3), 63-77.
8. Eidelman, A. I., Breastfeeding and the use of human milk: an analysis of the American Academy of Pediatrics 2012 Breastfeeding Policy Statement. *Breastfeed Med* **2012,** *7* (5), 323-4.
9. Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X., Recent and potential developments of biofluid analyses in metabolomics. *J Proteomics* **2012,** *75* (4), 1079-88.
10. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J Proteome Res* **2007,** *6* (2), 443-58.
11. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc* **2010,** *5* (6), 1005-18.
12. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2007,** *2* (11), 2692-703.

13. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Consortium, H. S. M., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **2011,** *6* (7), 1060-83.
14. Marincola, F. C.; Noto, A.; Caboni, P.; Reali, A.; Barberini, L.; Lussu, M.; Murgia, F.; Santoru, M. L.; Atzori, L.; Fanos, V., A metabolomic study of preterm human and formula milk by high resolution NMR and GC/MS analysis: preliminary results. *J Matern Fetal Neonatal Med* **2012,** *25* (Suppl 5), 62-7.
15. Pratico, G.; Capuani, G.; Tomassini, A.; Baldassarre, M. E.; Delfini, M.; Miccheli, A., Exploring human breast milk composition by NMR-based metabolomics. *Natural product research* **2014,** *28* (2), 95-101.
16. Giuffrida, F.; Cruz-Hernandez, C.; Flück, B.; Tavazzi, I.; Thakkar, S. K.; Destaillats, F.; Braun, M., Quantification of phospholipids classes in human milk. *Lipids* **2013,** *48* (10), 1051-8.
17. Daud, A. Z.; Mohd-Esa, N.; Azlan, A.; Chan, Y. M., The trans fatty acid content in human milk and its association with maternal diet among lactating mothers in Malaysia. *Asia Pac J Clin Nutr* **2013,** *22* (3), 431-42.
18. Kelishadi, R.; Hadi, B.; Iranpour, R.; Khosravi-Darani, K.; Mirmoghtadaee, P.; Farajian, S.; Poursafa, P., A study on lipid content and fatty acid of breast milk and its association with mother's diet composition. *J Res Med Sci* **2012,** *17* (9), 824-7.
19. Zou, X.; Huang, J.; Jin, Q.; Guo, Z.; Liu, Y.; Cheong, L.; Xu, X.; Wang, X., Lipid Composition Analysis of Milk Fats from Different Mammalian Species: Potential for Use as Human Milk Fat Substitutes. *J Agric Food Chem* **2013**.
20. Miller, J. B.; Bull, S.; Miller, J.; McVeagh, P., The oligosaccharide composition of human milk: temporal and individual variations in monosaccharide components. *Journal of pediatric gastroenterology and nutrition* **1994,** *19* (4), 371-6.
21. Blaas, N.; Schüürmann, C.; Bartke, N.; Stahl, B.; Humpf, H. U., Structural profiling and quantification of sphingomyelin in human breast milk by HPLC-MS/MS. *J Agric Food Chem* **2011,** *59* (11), 6018-24.
22. Garcia, A.; Barbas, C., Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods Mol Biol* **2011,** *708*, 191-204.

23. Whiley, L.; Godzien, J.; Ruperez, F. J.; Legido-Quigley, C.; Barbas, C., In-vial dual extraction for direct LC-MS analysis of plasma for comprehensive and highly reproducible metabolic fingerprinting. *Anal Chem* **2012,** *84* (14), 5992-9.
24. Trygg, J.; Holmes, E.; Lundstedt, T., Chemometrics in metabonomics. *J Proteome Res* **2007,** *6* (2), 469-79.
25. Fisher, R. A., *Statistical methods for research workers.* 7th ed.; Edinburgh Oliver and Boyd 1938: 1954; p 356.
26. Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K.; Fraser, P. D.; Bramley, P. M., Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J Exp Bot* **2005,** *56* (410), 219-43.
27. Zhang, J.; Koo, I.; Wang, B.; Gao, Q. W.; Zheng, C. H.; Zhang, X., A large scale test dataset to determine optimal retention index threshold based on three mass spectral similarity measures. *J Chromatogr A* **2012,** *1251*, 188-93.
28. Naz, S.; Garcia, A.; Barbas, C., Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Anal Chem* **2013,** *85* (22), 10941-8.
29. Sandra, K.; Pereira, A. o. S.; Vanhoenacker, G.; David, F.; Sandra, P., Comprehensive blood plasma lipidomics by liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Chromatogr A* **2010,** *1217* (25), 4087-99.
30. Tijerina-Sáenz, A.; Innis, S. M.; Kitts, D. D., Antioxidant capacity of human milk and its association with vitamins A and E and fatty acid composition. *Acta Paediatr* **2009,** *98* (11), 1793-8.
31. Canfield, L. M.; Clandinin, M. T.; Davies, D. P.; Fernandez, M. C.; Jackson, J.; Hawkes, J.; Goldman, W. J.; Pramuk, K.; Reyes, H.; Sablan, B.; Sonobe, T.; Bo, X., Multinational study of major breast milk carotenoids of healthy mothers. *Eur J Nutr* **2003,** *42* (3), 133-41.
32. Crill, C. M.; Helms, R. A., The use of carnitine in pediatric nutrition. *Nutr Clin Pract* **2007,** *22* (2), 204-13.
33. Martysiak-Żurowska, D., Content of odd-numbered carbon fatty acids in the milk of lactating women and in infant formula and follow-up formula. *ACTA Scientiarum Polonorum Technologia Alimentaria* **2008,** *7* (2), 8.
34. Posma, J. M.; Robinette, S. L.; Holmes, E.; Nicholson, J. K., MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics* **2014,** *30* (6), 893-5.

35. Bitman, J.; Wood, D. L.; Mehta, N. R.; Hamosh, P.; Hamosh, M., Comparison of the cholesteryl ester composition of human milk from preterm and term mothers. *Journal of pediatric gastroenterology and nutrition* **1986,** *5* (5), 780-6.
36. Jackson, M. B.; Lammi-Keefe, C. J.; Jensen, R. G.; Couch, S. C.; Ferris, A. M., Total lipid and fatty acid composition of milk from women with and without insulin-dependent diabetes mellitus. *Am J Clin Nutr* **1994,** *60* (3), 353-61.

Supporting Information (SI)

Authors: (CEMBIO) Alma Villaseñor[1]; Antonia García[1]; Coral Barbas[1]

(IMPERIAL) Isabel Garcia-Perez[1,2], Joram M. Posma[1], Nicholas Andreas [1], Elaine Holmes[1].

Address: [1]CEMBIO (Center for Metabolomics and Bioanalysis), Facultad de Farmacia, Universidad San Pablo CEU, Madrid, Spain.

[1]Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ London, UK.

[2]Nutrition and Dietetic Research Group, Division of Endocrinology and Metabolism, Imperial College London, W12 0NN, UK

Title: Breast milk metabolome characterization in a single phase extraction, multiplatform analytical approach.

Table of contents:

Figure S1. Extracted Ion chromatogram (EIC) in positive mode of some common fatty acids (FA) founded in BM. Ions were extracted by Find by Formula algorithm (Mass Hunter-Agilent®) at 10 ppm mass accuracy from the formula and with a match score was above 80 (MFG score). Highlighted peaks were selected according to the retention time of each standard.

Figure S-2. (A) PCA scores plot of the first 3 components of breast milk profiling obtained by LC-QTOF-MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (B) PCA pairs plot of the first 3 components of breast milk profiling obtained by GC–Q–MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (C) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by LC-QTOF-MS (D) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by GC–Q–MS.

Reagents.

LC-QTOF-MS. Analytical grade ammonia hydroxide (30% ammonium in high purity water) was acquired from Panreac Quimica SA (Barcelona, Spain) and analytical grade methyl-*tert*-butyl-ether (MTBE) from Sigma- Aldrich (Steinheim, Germany). LC-MS grade methanol, acetonitrile and analytical grade formic acid were purchased from Fluka Analytical (Sigma- Aldrich, Steinheim, Germany). Ultrapure water was obtained from a Milli-Qplus185 system (Millipore, Billerica, MA, USA). Vitamin-E-acetate (DL-α-Tocopheryl acetate) with a monoisotopic mass of 472.3931 ($C_{31}H_{52}O_3$), was purchased from Fluka Biochemica (Switzerland) and was used as internal standard (SI).

GC-Q-MS. Standards and reagents were all of analytical grade except where stated otherwise. C18:0 methyl ester (IS), *O*-Methylhydroxylamine hydrochloride, pyridine (silylation grade), heptane, acetone, isopropanol, methanol, and acetonitrile (last three LC-MS grade) were obtained from Sigma-Aldrich (Steinheim, Germany). N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) + 1% trimethylchlorosilane (TMCS) were purchased from Pierce Chemical Co (Rockford, IL, USA). Stock reference mix solution (Grain FAME mix) was purchased from Supelco/ Sigma-Aldrich (Bellefonte, USA). Standard solution was composed of a mix of 19 fatty acid methyl esters prepared in dichloromethane [10mg/ml]. This solution was kept at -20 °C and was diluted 1/100 before GC-MS analysis.

Table S-1. List of compounds chosen for validation for LC-QTOF-MS and GC-Q-MS.

No.	Compound	Biochemical class	Log P*	RT (min)	LC-QTOF-MS	GC-Q-MS
1	**Carnitine**	Alkylamines	-4.52±0.73	0.88	+	-
2	**α-Linolenic acid**	Fatty Acids	6.50±0.33	4.66	+	-
3	**Stearic acid**	Fatty Acids	8.22±0.19	9.20 (LC) 20.63 (GC)	+	+
4	**Cholesterol**	Steroids and derivatives	9.85±0.28	13.48	+	-
5	**Vitamin E**	Prenol Lipids	11.90±0.27	15.85	+	-
6	**Palmitic Sphingosine**	Sphingolipids	13.34±0.57	17.75	+	-
7	**CE 18:2**	Steroids and derivatives	18.17	29.60	+	-
8	**TG 18:2**	Triradylglycerols	22.16±0.44	30.07	+	-
9	**Lactic acid**	Hydroxy Acids	-0.70±0.27	6.78	-	+
10	**Urea**	Ureas	-2.11±0.19	9.38	-	+
11	**Myoinositol**	Cyclic Alcohols	-2.11±0.49	19.25	-	+
12	**Proline**	Amino Acids	-0.57±0.28	10.17	-	+
13	**Glutamic**	Amino Acids	-1.67±0.47	14.29	-	+
14	**Serine**	Amino Acids	-1.58±0.33	9.93	-	+
15	**Glucose**	Monosaccharides	-3.17±0.86	17.37	-	+
16	**Citric acid**	Carboxylic Acids	-1.72±0.40	16.52	-	+
17	**Glycerol**	Sugar Alcohols	-2.32±0.49	9.83	-	+

+/- denotesnwhether the compound was validated (+) or not (-) by each technique. * Values obtained from: chemspider website (http://www.chemspider.com/).

Table S-2. Tentative characterization of compounds present in breast milk profile by LC-QTOF-MS in positive and negative mode.

Category	Subcategory	Positive Mode LC-MS	Negative Mode LC-MS	Shared
Fatty Acyls	**Fatty Acids & Conjugates**	22	33	6
	Fatty alcohols	10	1	0
	Fatty aldehydes	11	4	1
	Fatty amides	3	0	0
	Fatty esters	8	16	5
	Octadecanoids	2	1	0
	Eicosanoids	0	3	0
	Other Fatty Acyls	0	1	0
Glycerolipids	**Monoacylglycerols**	41	0	0
	Diacylglycerols	48	0	0
	Triacylglycerols	22	0	0
	Ladderans	9	0	0
Glycerophospholipids	**Glycerophosphates**	8	5	0
	Glycerophosphocholines	22	5	3
	Glycerophosphoethanolamines	24	9	6
	Glycerophosphoglycerols	2	3	0
	Glycerophosphoinositols	1	1	0
	Glycerophosphoserines	1	1	0
	Glycosylglycerophospholipids	0	1	0
Sphingolipids	**Phosphosphingolipids**	23	0	0
	Ceramides	1	0	0
	Phosphonosphingolipids	1	0	0
	Glycosphingolipids	2	0	0
	Sphingoid bases	0	1	0
	Neutral glycolsphingolipids	0	1	0
Prenol Lipids	**Quinones &hydroquinones**	2	0	0
	Isoprenoids	3	1	0
Sterol Lipids	**Bile acids and derivatives**	5	0	0
	Secosteroids	10	4	0
	Steroids	4	8	2
	Sterols	2	0	0
Carbohydrates	**Monosaccharides**	0	2	0
	Oligosaccharides	0	3	0
	Sugar Acids	0	1	0

	Sugar Alcohols	0	1	0
	Trisaccharides	0	2	0
Others	**Tripeptide**	0	13	0
	Modified dipeptide		2	
	Cofactors	0	1	0
	Organic acids	0	1	0
	Carnitines	1	0	0
	Amino Acids	1	1	0

Tentative identification was made via our website http://biolab.uspceu.com/mediator/public which search the neutral mass into three databases (METLIN, Lipid maps and KEGG). Hits pertaining to secondary plant metabolites, pharmaceutical or synthetic compounds were discarded.

Table S-3. Compounds identified in breast milk profile by GC-Q-MS.

No.	Compound	Target Ion (Da)	Retention time (min)	No.	Compound	Target Ion	Retention time (min)
1	Pyruvic acid	174	6.76	**31**	Lyxose 2 / Lyxosylamine 2/ Ribose	103	15.06
2	Lactic acid (Standard confirmed)	147	6.92	**32**	Xylitol	217	15.47
3	Glycolic acid	147	7.13	**33**	Fucose 1	117	15.59
4	Valine 1	72	7.43	**34**	Fucose 2	117	15.73
5	Alanine 1	116	7.53	**35**	Citric acid (Standard confirmed)	273	16.57
6	2-Hydroxybutyric acid	147	7.87	**36**	Hippuric acid 2	105	16.88
7	2-Furoic acid	125	7.99	**37**	Myristic acid	117	16.91
8	Isoleucine 1	86	8.57	**38**	Tagatose 1 / Sorbose 2 / Sorbose 1 / Fructose 1	103	17.12
9	Valine 2	144	9.14	**39**	Tagatose 2 / Fructose 2 / Fructose 1	103	17.24
10	Urea (Standard confirmed)	147	9.54	**40**	Galactose 1 / Mannose 1 / Allose 1 / Gluconic acid lactone 1	205	17.28
11	Benzoic acid	179	9.62	**41**	Glucose 1 / Talose 1 (Standard confirmed)	319	17.40
12	Caprylic acid	201	9.81	**42**	Altrose 2 / Mannose 2 / Glucose 2 / Allose 2 / Talose 2	319	17.54
13	Glycerol (Standard confirmed)	147	9.92	**43**	Mannitol / Sorbitol	319	17.87
14	Phosphoric acid	299	10.00	**44**	1-Hexadecanol	299	17.95
15	Proline 2 (Standard confirmed)	142	10.28	**45**	Palmitoleic acid	311	18.68
16	Glycine	174	10.39	**46**	Palmitic acid	117	18.88
17	Succinic acid	147	10.48	**47**	N-acetyl-D-mannosamine 1 / N-acetyl-D-mannosamine 2	319	19.19
18	Glyceric acid	189	10.74	**48**	Myo-inositol (Standard confirmed)	318	19.32
19	Serine 2 (Standard confirmed)	204	11.14	**49**	Methyl Stearate (Internal standard)	74	19.66
20	Threonine 2	218	11.48	**50**	Heptadecanoic acid	327	19.81
21	Capric acid	229	12.39	**51**	Linoleic acid	75	20.42
22	Malic acid	233	12.80	**52**	Oleic acid	339	20.48
23	Adipic acid	111	13.00	**53**	Stearic acid (Standard confirmed)	341	20.69
24	Threitol	217	13.06	**54**	Arachidic acid	369	22.37
25	Pyroglutamic acid	156	13.21	**55**	Lactulose 1	204	23.92
26	Glutamic acid 1	174	13.33	**56**	Sucrose	361	24.10
27	Creatinine	115	13.62	**57**	Lactose 1 / Cellobiose 1	361	24.46
28	Glutamic acid 2 (Standard confirmed)	246	14.37	**58**	Trehalose / Maltose 1 / Maltose 2	361	24.91
29	Lauric acid	257	14.75	**59**	Galactinol 1	204	26.30
30	Lyxose 1 / Lyxosylamine 1	103	14.89	**60**	Cholesterol	129	27.64

Compounds were assigned using the Fiehn RTL and NIST08 libraries or by addition of the corresponding standard. For conditions see text. Numbers 1or 2 refer to the number of trimethylsilyl groups in the molecule after derivatization.

Table S-4. Validation parameters for selected metabolites in BM by LC-QTOF-MS.

Validation criteria	Carnitine	Stearic acid	Vitamin E	Palmitic Sphingosine	CE 18:2	Cholesterol	α-Linolenic acid	TG 18:2
Linearity								
Slope	6.17E+05 ± 2.63E+04	4.98E+03 ± 4.14E+02	1.49E+05 ± 5.39E+03	1.27E+06 ± 8.19E+04	2.79E+05 ± 2.83E+04	4.04E+04 ± 3.97E+03	8.92E+03 ± 7.22E+02	1.25E+06 ± 9.03E+04
Intercept	(-)1.36E+04 ± 3.31E+04	(-)6.55E+02 ± 9.77E+03	(-)8.05E+03 ± 9.33E+03	5.78E+03 ± 3.87E+03	(-)1.38E+04 ± 1.72E+04	3.12E+05 ± 1.49E+05	9.79E+03 ± 2.15E+04	4.42E+05 ± 9.17E+05
R	0.998	0.991	0.998	0.995	0.992	0.990	0.993	0.995
Range (ppm)	0.24 - 1.93	4.79 - 38.33	0.35 - 2.82	0.015 - 0.07	0.15 - 1.21	9.95 - 59.68	5.89 - 47.13	1.95 - 15.60
Recovery								
(%)	63.8	102.5	89.4	98.2	51.5	99.0	81.6	82.0
RSD (%)	11.0	5.3	8.1	6.3	15.0	6.8	7.98	4.68
Instrumental precision (n=10), % RSD	3.6	4.4	5.4	2.8	4.8	2.9	3.5	7.0
Method precision with standard								
Intra-day (n=7), % RSD	1.9	5.9	1.4	2.5	4.3	2.3	3.4	2.1
Inter-day (n=14), % RSD	1.9	5.0	2.1	2.8	3.3	3.9	3.6	2.3
Method precision with sample								
Intra-day (n=7), % RSD	2.6	5.8	3.0	3.2	2.6	1.4	4.5	4.1
Inter-day (n=14), % RSD	4.3	5.8	5.3	3.9	5.7	6.6	5.4	8.2
LOD (ppm)	0.016	1.168	0.065	0.004	0.016	0.225	0.743	0.006
LOQ (ppm)	0.053	3.895	0.217	0.015	0.053	0.749	2.475	0.023
Concentration in the BM pool sample [ppm ± (SD)]								
	0.86 ± (0.04)	21.00 ± (1.22)	1.49 ± (0.08)	1.84 ± (0.07)	0.42 ± (0.02)	1.24 ± (0.08)	0.88 ± (0.05)	1.28 ± (0.11)

Table S-5. Validation parameters for selected metabolites in BM by GC-Q-MS.

Validation criteria	Lactic acid	Urea	Glycerol	Myoinositol	Stearic acid	Proline	Glutamic	Glucose	Citric acid	Serine
Linearity										
Slope	9.17E-02 ± 5.83E-03	4.03E-02 ± 3.77E-03	1.62E-02 ± 1.58E-03	5.85E-02 ± 3.90E-03	4.37E-02 ± 2.90E-03	9.02E-02 ± 7.67E-03	2.19E-02 ± 1.66E-03	3.10E-02 ± 2.05E-03	2.25E-02 ± 1.51E-03	3.64E-02 ± 2.25E-03
Intercept	1.27E-01 ± 7.15E-02	(-)1.36E+00 ±7.24E-01	(-)3.59E-01 ± 4.94E-01	(-)2.82E-01 ± 6.19E-01	3.25E-01 ± 9.67E-02	(-)2.94E-02 ± 3.17E-02	(-)1.47E-02 ± 5.33E-02	(-)1.30E-01 ± 1.97E-01	(-) 1.66E-02 ± 2.51E-02	1.68E-02 ± 1.51E-02
R	0.995	0.991	0.992	0.994	0.994	0.990	0.992	0.995	0.994	0.997
Recovery										
Range (ppm)	2.89 - 23.09	39.03 - 312.20	70.40 - 563.20	32.32 - 258.56	6.55 - 52.42	0.84 - 6.72	6.54 - 52.32	18.7 - 149.64	2.89 - 23.09	1.88 - 15.04
Recovery (%)	97.7	106.1	98.3	95.7	90.7	82.5	96.5	90.5	80.2	104.1
RSD (%)	1.8	5.1	13.1	3.0	4.0	8.0	8.0	8.0	6.9	18.6
Instrumental precision (n=10), % RSD	2.8	3.6	3.0	3.3	5.5	7.7	5.8	3.9	3.3	6.9
Method precision with standard										
Intra-day (n=7), % RSD	3.5	5.8	5.1	3.2	2.8	7.9	3.3	3.7	3.5	6.3
Inter-day (n=14), % RSD	4.5	5.3	7.8	3.6	3.4	8.6	6.9	4.1	3.6	7.9
Method precision with sample										
Intra-day (n=7), % RSD	3.5	4.1	4.0	3.8	6.5	6.8	3.0	4.7	5.8	3.9
Inter-day (n=14), % RSD	5.0	4.7	5.2	3.9	8.6	10.8	6.6	6.7	7.8	4.7
LOD (ppm)	0.562	3.216	4.632	1.93	0.394	0.133	1.510	0.531	0.204	0.696
LOQ (ppm)	1.874	10.721	15.439	6.44	1.313	0.400	5.031	1.594	0.612	2.321
Concentration in the BM pool sample [ppm ± (SD)]										
	11.96 ± (0.59)	179.68 ± (8.40)	446.66 ± (23.09)	146.44 ± (5.77)	20.02 ± (1.73)	2.49 ± (0.27)	99.58 ± (8.36)	106.78 ± (7.13)	72.92 ± (5.72)	4.73 ± (0.22)

Table S-6. Tentative identification of significant metabolites in BM between the first week (0-7 days) against post 4 weeks postpartum by LC-QTOF-MS.

1st week vs 4 weeks					
↑4 weeks			↓4 weeks		
Compound	Measured mass (Da)	Change (%)	Compound	Measured mass (Da)	Change (%)
LPE(18:2)	**477.2841**	58.51	Hydroxyadipic acid	**162.0528**	-64.98
Octadecatrienal / Ladderane-hexanol	**262.2284**	52.15	α-Tocopherol*	**430.3789**	-157.91
Oleic acid C18:1*	**282.2551**	124.96	Cholesterol*	**369.3508**	-24.36
Palmitoleic acid C16:1*	**254.2216**	72.59	CE 18:2*	**648.5830**	-69.55
Linoleic acid C18:2*	**280.2391**	110.98	LPC(16:0) / LPE(19:0)	**495.3312**	-80.02
Anandamide (20:4)	**347.2815**	57.96	LPC(15:0) / LPE(18:0)	**481.3154**	-90.94
Sphingosine / Ketosphinganine / Palmitoyl Ethanolamide/ Amino-octadecanoic acid	**299.2815**	38.21	PE(O-41:5)	**723.5184**	-70.09
Pregnanetriol	**336.2652**	59.13	PE(40:1) / PC(40:1)	**731.5438**	-75.63
SM(d41:1) / N-Stearoyl-D-sphingomyelin / PE-Cer(d41:1)	**730.5966**	42.72	PC(40:0) / PE(40:0)	**733.5608**	-110.70
Keto-tridecanoic acid / Methoxy-dodecenoic acid	**228.1717**	69.01	PC(42:1) / PE(42:1)	**759.5763**	-71.68
Keto-stearic acid / Hydroxy-octadecenoic acid	**298.2500**	50.66	PC(42:0) / PE(42:0)	**761.5920**	-63.61
Oxo-nonadecanoic acid	**312.2657**	76.52	PC(44:1) / PE(44:1)	**787.6067**	-86.06
Hydroxy-eicosenoic acid / Keto-eicosanoic acid	**326.2811**	46.63	SM(d39:1) / PE-Cer(d39:1)	**702.5654**	-60.26
MG(14:0)	**302.2453**	42.84	GlcCer(d48:1) / Galbeta-Cer(d48:1)	**811.6870**	-65.77
MG(15:0) / Dihydroxy stearic acid	**316.2603**	51.11	TG(61:2)	**942.8581**	-280.21
MG(16:0)	**330.2772**	44.98			
MG(18:2)	**354.2762**	52.27			
MG(18:1) / Heneicosanedioic acid	**356.2918**	53.50			
MG(18:0)	**358.3076**	47.58			
Hexadecyl-arachidonoyl-glycerol / Methyl-pentadecanoyl-ladderane-octanyl-sn-glycerol	**602.5253**	56.99			
DG(31:2)	**508.4114**	58.82			
DG(31:1)	**510.4277**	67.39			
DG(33:1)	**538.4600**	50.19			
DG(P-35:1)	**550.4944**	58.16			

DG(35:2)	**564.4755**	66.25			
DG(35:1)	**566.4916**	52.68			
DG(36:3)	**576.4735**	59.05			
DG(36:1)	**580.5055**	60.24			
DG(36:0)	**582.5205**	56.85			
DG(37:3)	**590.4905**	53.32			
DG(37:2)	**592.5079**	54.94			
DG(37:0)	**596.5384**	56.06			
DG(39:3)	**618.5228**	52.37			
DG(39:2)	**620.5385**	50.68			
DG(39:1)	**622.5533**	44.83			
DG(43:7)	**666.5202**	76.28			
TG(41:1)	**664.5623**	63.31			
TG(43:2)	**690.5786**	62.37			
TG(48:2)	**760.6563**	53.89			
TG(52:2)	**816.7197**	33.55			
TG(54:2)	**844.7529**	31.22			

Tentative identification of metabolites significantly associated with time in breast milk composition. *Confirmed by standard. For putative identification, all the compounds presented an estimated formula score >90% matched with the experimental isotopic pattern distribution on Mass Hunter software. Change (%); +/- increased/decreased in BM with the time, this was calculated as follows: (Average [4 week]-Average [1 week])/Average [1 week]*100).

LPC, LPE and MG –have been named with the number of carbons in the fatty acid chain attached to the backbone and the number of unsaturated bonds, e.g., LPC (16:0). PE, PC, DG and TG- have been named with the total number of carbons in the molecule and the number of unsaturated bonds, e.g., PC(42:0).

Table S-7. Tentative identification of significant metabolites in BM between the first week (0-7 days) against >4 weeks postpartum by GC-Q-MS.

1st week vs 4 weeks					
↑4 weeks			↓4 weeks		
Compound	**Ref. *m/z***	**Change (%)**	**Compound**	**Ref. *m/z***	**Change (%)**
Oleic acid C18:1	**339**	**105.96**	Fucose	**117**	-102.49
Palmitoleic acid C16:1	**311**	**79.58**	Furanose isomers	**103**	-85.22
Linoleic acid C18:2	**75**	**97.62**	D-glucosaminic acid	**291**	-73.85
			Cholesterol	**129**	-134.92

Change (%); +/- increased/decreased in BM with the time, this was calculated as follows: (Average [4 week]-Average [1 week])/Average [1 week]*100).

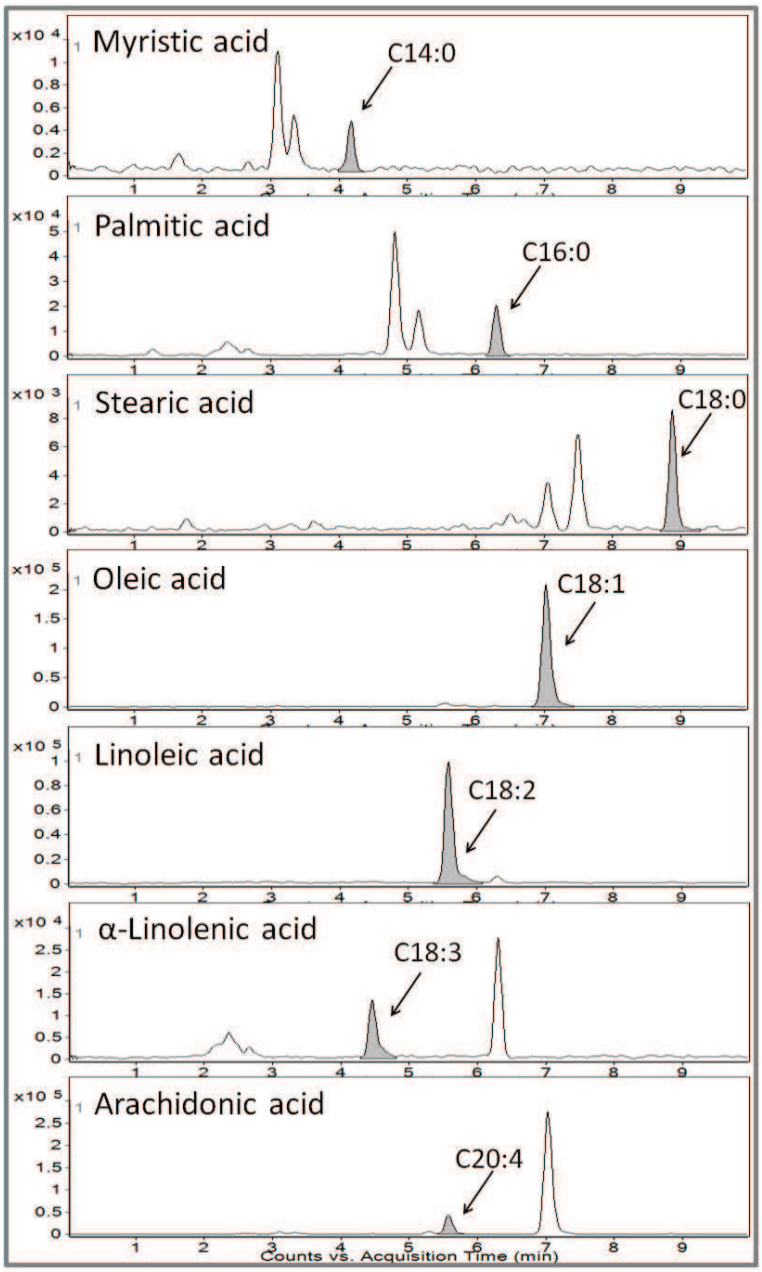


Figure S1. Extracted Ion chromatogram (EIC) in positive mode of some common fatty acids (FA) founded in BM. Ions were extracted by Find by Formula algorithm (Mass Hunter-Agilent®) at 10 ppm mass accuracy from the formula and with a match score was above 80 (MFG score). Highlighted peaks were selected according to the retention time of each standard.

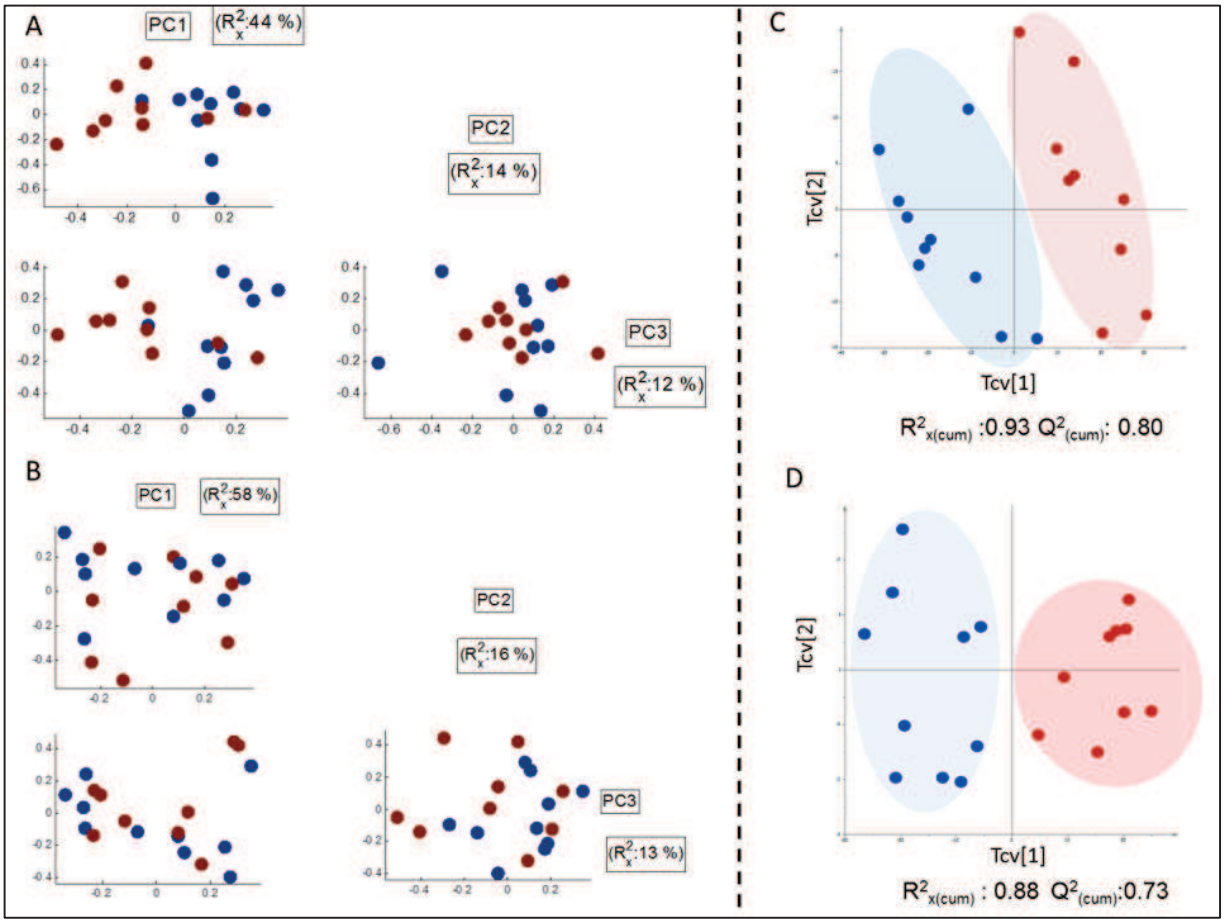
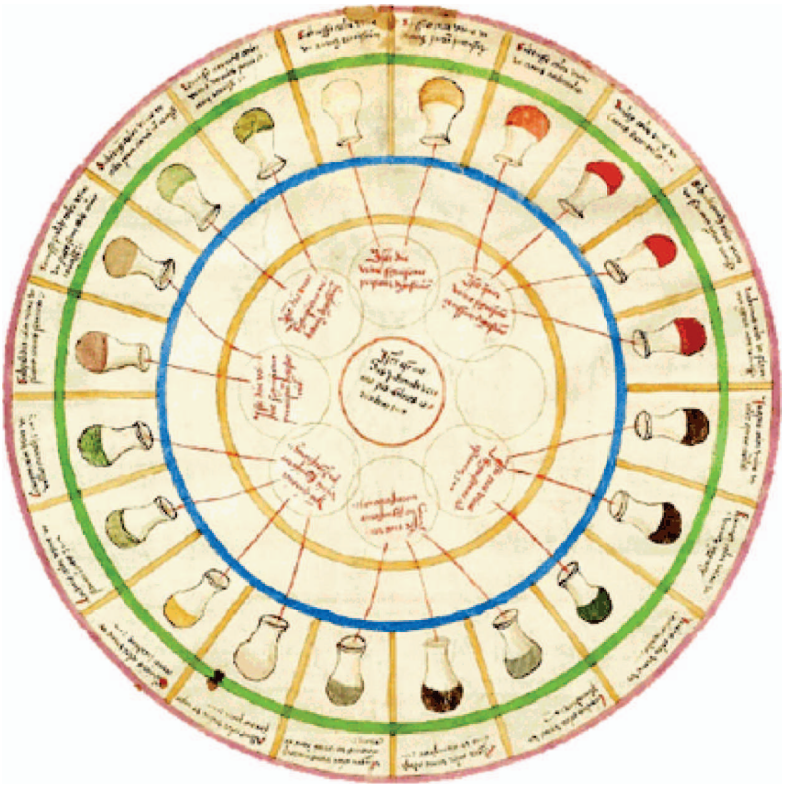


Figure S-2. (A) PCA scores plot of the first 3 components of breast milk profiling obtained by LC-QTOF-MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (B) PCA pairs plot of the first 3 components of breast milk profiling obtained by GC–Q–MS comparing samples collected on days 1-7 post-partum (blue dots) vs samples collected more than 4 weeks post-partum (red dots). (C) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by LC-QTOF-MS (D) PLS-DA scores plot of breast milk samples collected during the first 7 days post-partum (blue dots) vs breast milk samples collected after 4 weeks post-partum (red circles) obtained by GC–Q–MS.



SECOND SECTION.

METABOLIC NON–TARGET ANALYSIS

APPLICATIONS

CHAPTER 3.

2. Nuclear Magnetic Resonance (NMR) spectroscopy

3.1. Fundamentals of the technique

NMR spectroscopy functions by the application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms. When a sample is placed into a magnetic field and is subjected to RF radiation (energy) at the appropriate frequency, the nuclei in the sample can absorb the energy. After the absorption of energy by the nuclei, the length of time and the manner in which the nuclei emit this energy during the relaxation process is used to reveal information regarding a variety of dynamic processes. The frequency of the radiation necessary for absorption of energy will depend on three aspects: (*i*) the characteristic of the type of nucleus (e.g. ^{1}H or ^{13}C); (*ii*) the chemical environment of the nucleus, e.g. the methyl and hydroxyl protons from methanol absorb at different frequencies than the amide protons in a native protein since they are in different environments; (*iii*) the strength of the magnetic field, which should not vary over the sample (shimming)[1–2].

All nuclei have positive charges and many of them behave as though they were spinning, referred to as "spin" (I). Most nuclei of biological interest, including ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P, have a nuclear spin quantum number of I = l/2. This means that the nucleus acts as a miniature bar magnet. Anything that is charged and moves has a magnetic moment and produces a magnetic field. If we put this small magnet in the field of a much larger magnet its orientation will no longer be random, and moreover if RF waves are applied, the nuclei in the lower energy state can absorb the energy and jump to the higher energy state, in a "flip" movement, while the energy level is induced. We can observe either the absorption of energy, or the subsequent release of energy as the nucleus "relaxes" back to the lower energy state. The energy absorbed by nuclear spins induces a voltage that can be detected by a suitably tuned coil of wire, amplified and the signal can be displayed as *free induction decay (FID)*[2].

The energy required to induce flipping and to obtain an NMR signal is determined by the difference in energy levels between the two nuclear orientation states ΔE (lower or upper energy states),

as the nuclear spin could be oriented in one of these two states. The energy, *E*, is related to its frequency, ν, by Planck's constant (h)[2]:

$$\Delta E = h \times \nu$$

Here the quantity, ν, is called the resonance frequency, and is the angular *Larmor resonance frequency*. In practice, we do not have one nucleus in a magnetic field but rather a huge number; the distribution of nuclei in the different energy states under the application of any RF energy is given by Boltzmann equation[2].

$$\frac{Nupper}{Nlower} = e^{-\Delta E/kT} = e^{-h\nu/kT}$$

N_{upper} and N_{lower} represent the population, i.e. the number of nuclei in upper and lower energy states respectively, *k* is the Boltzmann constant, and T is the absolute temperature (˚K). The NMR signal is thus proportional to the population difference between the states, which is usually small. The low sensitivity of NMR has its origin here, and is probably its greatest limitation for applications to biological systems. As can be seen from the two equations above, the use of stronger magnetic fields (↑ν) will increase the population ratio (from Boltzmann equation) and consequently, the sensitivity[2].

Finally, the FID is a time–domain signal; the usual frequency–domain spectrum can be obtained by computing the Fourier transformation of the signal–averaged FID, and is usually presented in the form of 1D spectrum. The general features that describe a typical NMR spectrum are shown in Figure 3.1[2]:

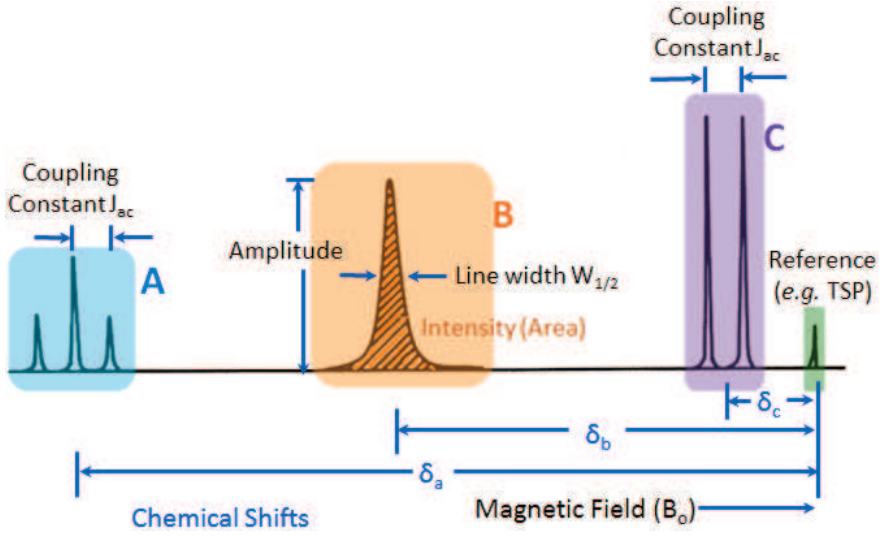


Figure 3.1. Nuclear magnetic resonance spectral parameters: chemical shifts, area, line width, splitting (taken from James, T. L., 1998[2]).

1. *Chemical shift.* The resulting resonance frequency will depend on the exact characteristics of the electronic environment around the nucleus. The induced magnetic fields are typically a million times smaller than the applied magnetic field. So if the Larmor resonance frequency, $\nu_{o,}$ is on the order of several megahertz, differences in resonance frequencies for two different hydrogen nuclei, for example, will be on the order of several hertz. Although we cannot easily determine absolute RF to an accuracy of ±1Hz, we can determine the relative positions of two signals in the NMR spectrum with greater accuracy. Consequently, a reference signal is chosen, and the difference between the position of the signal of interest and that of the reference is termed as the "*chemical shift*". This is a dimensionless number, expressed with the symbol δ and is usually measured in terms of parts per million (ppm)[2]. Protons will normally resonate in the region between 0 and 12 ppm. Typical ^{1}H NMR chemical shifts ranges are observed in Figure 3.2.
2. *Spin–Spin Coupling (Splitting)*. A nucleus with a magnetic moment may interact with other nuclear spins, resulting in mutual splitting of the NMR signal from each nucleus into multiplets. Multiplet splitting indicates the number of neighbouring hydrogen spin–spin coupling. The difference between any two adjacent components of a multiplet is the same, and yields the value of the spin–spin coupling constant *J* (in hertz)[2].
3. *Line width and peak intensity*. The line width can sometimes reveal details about molecular geometry. The area of an NMR signal is directly proportional to the number of nuclei contributing to the signal. Absolute concentrations can be obtained if the sample contains an added internal standard of known concentration[3].

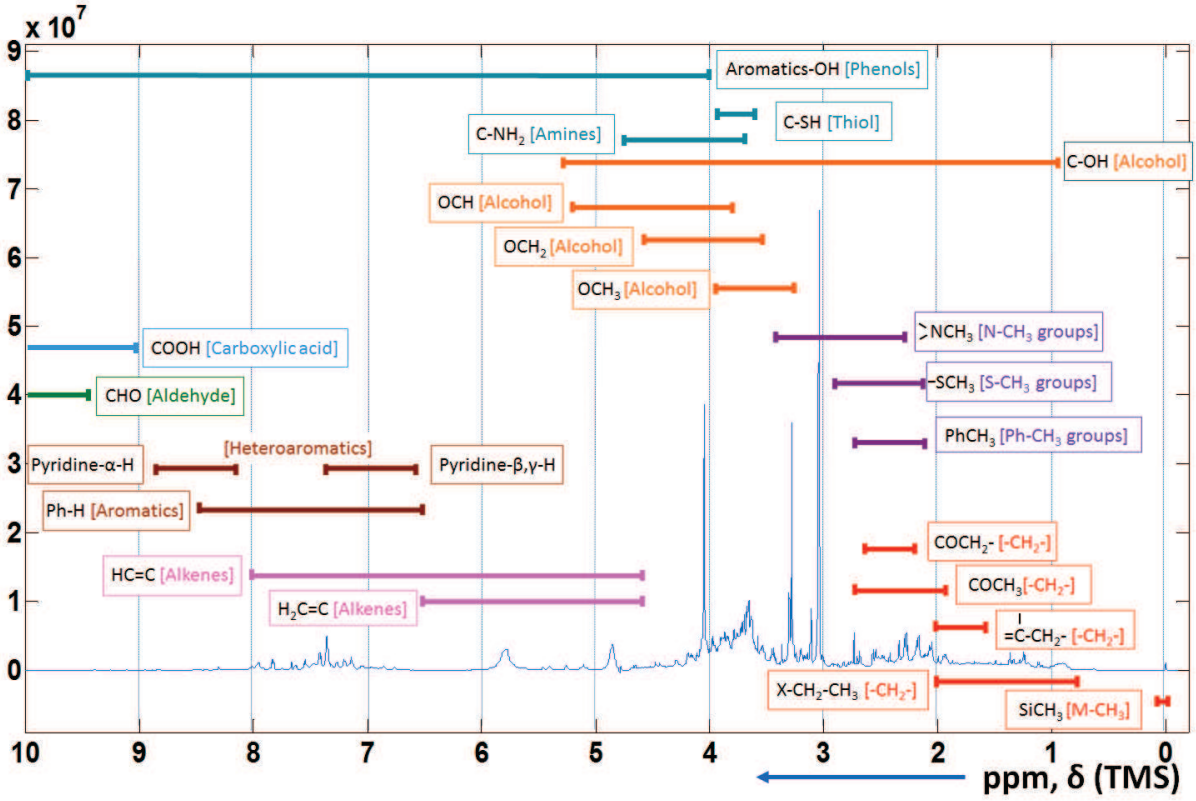


Figure 3.2. Typical ^{1}H NMR chemical shift ranges for organic compounds. The size of the shift in ppm is relative to the TMS peak which is referenced at 0 ppm. The urinary ^{1}H NMR spectrum was taken from a patient with diverticulitis. The chemical shifts values were taken from NMR Guide3.5, Copyright © 1998–2003 BRUKER Biospin. All rights reserved.http://triton.iqfr.csic.es/guide/man/beginners/chap3–4.htm.

Some of the general aspects that should be taken into account for NMR are: (*i*) the pH of the sample should be fixed by adding a buffer to prevent chemical shifting and to minimize differences in ionic strength[4]; (*ii*) a small amount of D_2O must be added, as this permits the frequency–field lock to offset the effect of natural drift of the NMR magnetic field, so that all scans are exactly co–registered (field lock); (*iii*) an internal chemical shift and quantitation standard should be used, such as 3–trimethylsilylpropanoate (TSP); (*iv*) the resonances produced are very sensitive to structure, so many compounds can be identified in spectral data simultaneously, and the fine structure of the signals in the frequency domain provides strong structural clues for the identification of unknown metabolites[3]. In

addition, preservative compounds such as borate and azide should be considered as additives, as they are highly effective antimicrobial agents for urine samples[5].

3.2.Characteristics of NMR spectroscopy for metabolomics studies

NMR was one of the first analytical techniques used in the field of metabolomics. Specifically, it was in the late 1960s with the development of Fourier transform NMR spectroscopy, and then in the 1970s with the introduction of superconducting magnets, which resulted in the increase of the sensitivity, that the first applications of NMR spectroscopy started in the metabolic profiling of biofluids and cells[6]. Ever since, NMR has been one of the main workhorses in metabolomics research.

Generally, NMR spectroscopy provides a rapid, quantitative, non–destructive, unbiased, robust and reproducible technology, with good resolution and high–throughput that requires minimal sample preparation[1, 3, 7–8]. NMR can measure up to 400 samples per 24hrs with flow–injection technology[8]. In addition, it can operate in a non-targeted way, i.e. any sufficiently abundant molecule containing the nuclei of interest will be detected. NMR spectroscopy is also an acclaimed technique for providing detailed structural information of small organic molecules and, as such, has enabled a large number of biofluid constituents to be identified and listed in databases such as HMDB (Human Metabolomic Data Base)[9]. In addition, NMR is a considerable utility in the direct detection for identification of drugs and their metabolites in biological samples[10]. Other NMR–visible nuclei such as ^{13}C, ^{31}P and ^{15}N can also be exploited to a lesser extent, and provides complementary information normally for identification purposes.

NMR as a metabolomics analytical tool has additional characteristics; this allows a comprehensive analysis of metabolites in a single run. Here, NMR can generate information–rich spectra that represent the metabolic snapshot in any biological sample, i.e. any resulting spectrum of a biofluid offers an overview of the metabolic state of the organism at the moment of sampling[10–12]. For a trained NMR–spectroscopist, the resulting spectrum offers a visual idea of the major metabolites, and sometimes even reveals the metabolic state of the patient. In a sense, we could say NMR spectroscopy is more closely a universal detector in which the sample can be analysed directly, and many kinds of small

metabolites can be measured as a snapshot picture. To date, ^{1}H NMR has become a well–established, exceptionally reproducible and quantitative method used in metabolomics studies[13]. In addition, using a technique called high–resolution ^{1}H magic angle spinning (MAS) NMR spectroscopy, it is possible to acquire NMR spectra of intact tissues with no sample pretreatment[3].

However, NMR suffers from both poor sensitivity and limits of detection[7, 12]. Additionally, all biofluids are highly complex mixtures of metabolites, and even 1D spectra may contain many thousands of resolved NMR resonance lines with varying degrees of signal overlap, depending on the fluid type and the chemical shift range. Despite the powerful arsenal of multipulse and multidimensional NMR tools available, the complexity of the spectra and the degree of signal overlap has hindered the comprehensive assignment of the ^{1}H NMR spectra in most biofluids[14]. The assignment problems vary considerably between biofluid types, and are particularly hard for those metabolites present close to the limits of detection[15]. Also, in complex biofluids like plasma, macromolecules can give rise to the presence of broad resonances that can overshadow small molecules, making detection and quantitation difficult[7]. Thus, spin–eco sequence Carr–Purcell–Meiboom–Gill (CPMG) can be used to select only the contributions from the small–molecule metabolites[3]. These advantages and drawbacks of NMR spectrometry are summarized in Table 3.1.

In order to assign identification in ^{1}H NMR spectra, the addition of the standard to the sample is only one way. Additional confirmation can be done by the application of 2D NMR methods: COSY and TOCSY; which provide ^{1}H–^{1}H spin–spin coupling connectivities, giving information on which hydrogen atoms in a molecule are close in chemical–bond terms. HMQC, HMBC and HSQC are inverse–detected heteronuclear correlation methods, and the application of 2D J–resolved (JRES) yields information on the multiplicity and coupling patterns of resonances[15–16]. Additionally, statistical tools have been developed for aiding the assignment. Statistical Total Correlation Spectroscopy (STOCSY) is an analysis method that was developed by Cloarec, O in 2005[17]. This allows identification of highly correlated peak intensities leading directly to confirm the peaks from the same molecule and also the correlation of compounds in the same metabolic pathway where concentrations are dependent or under some common regulatory mechanism[17]. Examples of STOCSY can be found along this chapter.

Table 3.1 Relative strengths and weakness of NMR for metabolomics studies.

Strengths	Drawbacks
▪ Rapid, quantitative, non–destructive, unbiased, robust and reproducible, with good resolution	▪ Low sensitivity and limits of detection
▪ High–throughput technology (up to 400 samples per 24 h)	▪ Poor resolution for low concentrated metabolites in complex samples
▪ Minimum sample preparation	▪ Difficult to unequivocally assign identification
▪ Provide detailed structural information	▪ High spectra complexity and signal overlap
▪ Identification, from databases of standards (such as HMDB) and by self–consistent analysis of 1D and 2D spectra	▪ Macromolecules can give rise to broad resonances (like 1D–plasma)
▪ Suitable for liquids and solids	▪ Normally 2D experiments demand long time
▪ Whole sample analysed in one measurement	▪

3.3. Insight into the application of NMR

For the organization of this work, the application of NMR was chosen as the first paper in this section, because it was historically the first high–throughput technology used in metabolomics, and has subsequently become a reference and basic tool for researchers. Thus, its use has been expanded in multiple applications: toxicity, disease, monitoring therapeutic targets, natural products, functional genomics, etc.[18]. The next paper embraces a very ambitious objective: the application of NMR spectroscopy in the clinic. Normally when an experiment is planned, everything attempt is made to reduce variability between inter–groups and inside the study group, because the interpretation of the results is made more difficult and mistakes are more probable. Nevertheless, as it was set in the introduction section, one of the final steps of metabolomics is to be able to apply this technology to the human being and their health. This is not an easy task, and this can be reflected in the paper. The samples analysed here were from patients at the emergency setting in the hospital, where two groups of populations were chosen: inpatients with acute pancreatitis (AP); and inpatients with other types of diagnosis. Both characterized only by acute abdominal pain.

There were several analytical challenges for NMR in this case–study: the heterogeneity of the patients in the study was reflected in the spectra obtained in biological matrices for both urine and plasma. As it was commented previously, the visual inspection of each spectrum gave some idea of the physiological condition of the patients, e.g. the presence of ethanol and their metabolites, drugs such as acetaminophen, high peaks of ketone bodies, and so on. These aspects are expected from samples of patients like these, but this is not good for data treatment, as they can obscure more subtle endogenous changes resulting from the real condition of study. And this is exactly what was observed: a group of patients that were characterized by the presence of mannitol (a common excipient in intravenous preparations) were separated from the patients with higher and lower amounts of acetaminophen. Another similar example was reported by Maher, A. D. in 2012[19] for glucose in patients with diabetics. Although glucose is endogenous, this marked response is already well characterized in diabetes, and NMR peaks of glucose often obscure other peaks in the same chemical shift region[19]. In the case of AP, there is not a specific biomarker for diagnosis and this was one of the objectives of the study. Therefore in order to sort out the presence of non–related compounds from the study, these signals were removed from the raw data. Ideally all treatment-related metabolites should be removed, but even if this was feasible, the chances of unaware compounds are still possible. When xenobiotics were removed from the spectra (i.e. mannitol and acetaminophen with its related metabolites), the cut meant losing from the statistical analysis all other metabolites that overlapped the same chemical shift ranges.

Moreover, in the case of some spectra the presence of very significant broadened signals in most of the profile was observed. This can be due to excess proteinuria, making this kind of profile useless for the data treatment process i.e. in data alignment, so they were also excluded. Finally, when you work with multifactorial diseases like AP, the help of a medical support team is crucial in delimiting the classes of groups, and when to declare some samples are out of the scope of the project, e.g. chronic pancreatitis patients.

After the chemometric and multivariate analyses were performed, the potential statistically significant compounds were selected, so that the signals could be identified. For this particular study, most of the significant metabolites were very small so their confirmation was not easy, some of the unknowns were coming from drug related metabolites such as the case of antibiotic usage comparison. In addition, the 2D experiments were not always sensitive enough to resolve all the small compounds.

With all these challenges, and counting the small size of samples through 1D–NMR spectra of urine and plasma, we were able to show clear metabolic differentiation on three clinical sets of AP: (*i*) diagnosis of AP from other abdominal pain causes; (*ii*) characterization of principal aetiologies of the AP diseases (alcoholism vs cholelithiasis); and (*iii*) severity of the disease based on the clinical Glasgow score (mild AP vs severe AP). This approach was also able to reveal differences inside the control group, distinguishing cholelithiasis from colonic inflammation.

REFERENCES.

1. Dunn, W. B.; Ellis, D. I., Metabolomics: Current analytical platforms and methodologies. *Trends in Analytical Chemistry* **2005,***24* (4), 9.

2. James, T. L., Fundamentals of NMR. California., D. o. P. C. U. o., Ed. San Francisco, California 94143–0446 U.S.A., 1998; pp 1–31.

3. Lindon, J. C.; Nicholson, J. K., Analytical technologies for metabonomics and metabolomics, and multi–omic information recovery. *Trends in Analytical Chemistry* **2008,***27* (3), 194–204.

4. Lauridsen, M. B.; Bliddal, H.; Christensen, R.; Danneskiold–Samsøe, B.; Bennett, R.; Keun, H.; Lindon, J. C.; Nicholson, J. K.; Dorff, M. H.; Jaroszewski, J. W.; Hansen, S. H.; Cornett, C., ^{1}H NMR spectroscopy–based interventional metabolic phenotyping: a cohort study of rheumatoid arthritis patients. *J Proteome Res* **2010,***9* (9), 4545–53.

5. Smith, L. M.; Maher, A. D.; Want, E. J.; Elliott, P.; Stamler, J.; Hawkes, G. E.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Large–scale human metabolic phenotyping and molecular epidemiological studies via ^{1}H NMR spectroscopy of urine: investigation of borate preservation. *Anal Chem* **2009,***81* (12), 4847–56.

6. Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., NMR–based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem Res Toxicol* **2008,***21* (1), 9–27.

7. Price, K. E.; Lunte, C. E.; Larive, C. K., Development of tissue–targeted metabonomics. Part 1. Analytical considerations. *J Pharm Biomed Anal* **2008,***46* (4), 737–47.

8. Lao, Y. M.; Jiang, J. G.; Yan, L., Application of metabonomic analytical techniques in the modernization and toxicology research of traditional Chinese medicine. *Br J Pharmacol* **2009,***157* (7), 1128–41.

9. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J Proteome Res* **2007,***6* (2), 443–58.

10. Keun, H. C.; Athersuch, T. J., Nuclear magnetic resonance (NMR)–based metabolomics. *Methods Mol Biol* **2011,***708*, 321–34.

11. Rousseau, R.; Govaerts, B.; Verleysen, M.; Boulanger, B., Comparison of some chemometric tools for metabonomics biomarker identification. *Chemometrics and Intelligent Laboratory Systems.* **2008,***91* (1), 13.

12. Sandusky, P.; Raftery, D., Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey. *Anal Chem* **2005,***77* (8), 2455–63.

13. Sussulini, A.; Prando, A.; Maretto, D. A.; Poppi, R. J.; Tasic, L.; Banzato, C. E.; Arruda, M. A., Metabolic profiling of human blood serum from treated patients with bipolar disorder employing ^{1}H NMR spectroscopy and chemometrics. *Anal Chem* **2009,***81* (23), 9755–63.

14. Nicholson, J. K.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Lindon, J. C., 750 MHz ^{1}H and ^{1}H–^{13}C NMR spectroscopy of human blood plasma. *Anal Chem* **1995,***67* (5), 793–811.

15. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R., Metabonomics: Metabolic Processes Studied by NMR Spectroscopy of Biofluids. *Concepts in Magnetic Resonance* **2000,***12* (5), 289–320.

16. Lindon, J. C.; Nicholson, J. K., Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* **2008,***1*, 45–69.

17. Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J., Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Anal Chem* **2005,***77* (5), 1282–9.

18. Holmes, E.; Antti, H., Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analyst* **2002,***127* (12), 1549–1557.

19. Maher, A. D.; Fonville, J. M.; Coen, M.; Lindon, J. C.; Rae, C. D.; Nicholson, J. K., Statistical total correlation spectroscopy scaling for enhancement of metabolic information recovery in biological NMR spectra. *Anal Chem* **2012,***84* (2), 1083–91.

1H NMR global metabolic phenotpying of acute pancreatitis in the emergency unit.

Journal:	*Journal of Proteome Research*
Manuscript ID:	pr-2014-00161w
Manuscript Type:	Article
Date Submitted by the Author:	19-Feb-2014
Complete List of Authors:	Villaseñor, Alma; Universidad San Pablo CEU, Centre for Metabolomics and Bioanalysis (CEMBIO); Imperial College London, Department of Biomolecular Medicine Kinross, James; Imperial College London, Surgery; Imperial College London, Department of Biomolecular Medicine Li, Jia; Imperial College London, Department of Biomolecular Medicine Penney, Nicholas; Imperial College London, Surgery Barton, Richard; Imperial College London, Department of Biomolecular Medicine Nicholson, Jeremy; Imperial College London, Department of Biomolecular Medicine Darzi, Ara; Imperial College London, Surgery Barbas, Coral; Universidad San Pablo CEU, Centre for Metabolomics and Bioanalysis (CEMBIO) Holmes, Elaine; Imperial College London, Department of Biomolecular Medicine

^{1}H NMR global metabolic phenotpying of acute pancreatitis in the emergency unit.

Alma Villaseñor§, James M. Kinross†*, Jia V. Li*, Nicholas Penney†, Richard H. Barton*, Jeremy K. Nicholson*, Ara Darzi† Coral Barbas§ and Elaine Holmes*.*

† Section of Biosurgery & Surgical Technology, Division of Surgery, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, QEQM building, St. Mary's Hospital, London W2 1NY, UK.

* Section of Biomolecular Medicine, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2AZ, UK.

§Centre for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, Universidad San Pablo CEU, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

AUTHOR INFORMATION

Corresponding Authors

* Dr James M. Kinross,

Department of Biosurgery and Surgical Technology, QEQM building, St. Mary's Hospital, London, W2 1NY, UK.

Tel: +44 (0) 79 8934 4238

e-mail: J.kinross@imperial.ac.uk

* Prof. Elaine Holmes

Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ, UK.

Tel: +44 (0) 20 7594 3220

Fax: +44 (0) 20 7594 3226

e-mail: elaine.holmes@imperial.ac.uk

Keywords: Acute pancreatitis, Abdominal pain, Metabonomics, NMR, patient stratification.

ABSTRACT

We have investigated the urinary and plasma metabolic phenotype of acute pancreatitis patients presenting to the emergency room at St. Mary's Hospital London with acute abdominal pain in a single center, prospective study using ^{1}H NMR spectroscopy and multivariate modelling. Patients were allocated to either the Acute Pancreatitis (AP, n=15) or control group (all other causes of abdominal pain, n=21) on the basis of a diagnostic serum amylase measurement, and imaging (ultrasound or CT) as per national guidelines. Patients were assessed for three clinical outcome measures: 1) diagnosis of AP; 2) aetiology of AP caused by alcohol consumption and cholelithiasis; and 3) AP severity based on the Glasgow severity score. Samples from patients diagnosed with AP were characterized by high levels of urinary ketone bodies, glucose, plasma choline and lipid, and relatively low levels of urinary hippurate, creatine and plasma branched chain amino acids. AP could be reliably identified with a high degree of sensitivity and specificity (Q^2= 0.76 and R^2= 0.59) using panel of discriminatory urinary and plasma biomarkers for AP consisting of guanine, hippurate and creatine (urine) and valine, alanine and lipoproteins (plasma). High plasma levels of ketone bodies and 3-hydroxyisovalerate were associated with alcohol-induced AP. Higher plasma glucose concentrations were also associated with moderate severity of AP scored by Glasgow 2-3 relative to mild AP (Glasgow 1), particularly when accompanied by lower creatinine, alanine, valine, choline and acetylglycoprotein levels. Metabolic phenotyping was also able to distinguish between cholelithiasis and colonic inflammation amongst the heterogeneous control group to differentiate other common causes of abdominal pain associated with inflammation such as diverticulitis and appendicitis. Cholelithiasis patients demonstrated higher concentrations of plasma lactate, choline and lipid, and urinary indoleacetate while colonic inflammation patients had increased excretion of gut

microbial metabolites, malonate and 3-hydroxyisovalerate, and lower levels of valine, and leucine. This work has demonstrated that combinatorial metabolic biomarkers have a strong diagnostic and prognostic potential in AP with relevance to clinical decision making in the emergency unit.

INTRODUCTION

Acute Pancreatitis (AP) is an inflammatory condition associated with a progressive systemic inflammatory response and, in severe cases, autonecrosis of pancreatic tissue, organ failure and death(1, 2). In Europe, the mortality for this condition remains at 10%, and in severe cases associated with multiple organ failure it can be as high as 30%. AP affects between 5 and 80/100,000 people globally, and the main stay of therapy in the acute phase is the rapid initiation of supportive treatments, the treatment of reversible underlying causes and the allocation of intensive therapies, where needed. An early diagnosis is vital for initiating supporting therapy and minimizing morbidity, but this is often challenging. In the UK, an AP diagnosis is conventionally made on patient history, clinical examination and on the measurement of serum amylase. However, this lacks sensitivity and specificity, and it is subject to variation in its diagnostic threshold over time after the initial pancreatic insult(3). Although other biomarkers are available (e.g. serum pancreatic lipase, or urinary amylase), they do not provide prognostic value and they are expensive. Imaging modalities such as ultrasound (USS) and dynamic computed tomography (CT) scanning may also be used to confirm the diagnosis of AP(4); however USS is non-specific and a CT is only used to assess severity at 72 hours, or to rule out other diagnoses. Prognostication is essential in AP for early organ support and minimization of mortality. Unfortunately clinical risk scores such as the Glasgow score(5), are only marginally more accurate than clinical intuition(6-10). Other generic physiological scoring systems are used

in critical care such as the Acute Physiology And Chronic Health Evaluation (APACHE II) score(11), a generic physiological measurement based on 12 parameters, that is designed to measure the severity of disease for adult patients 24 hours following admission to intensive care units. However this is not disease specific for AP. Ultimately, diagnosing the aetiology of the underlying condition remains the only methodology for preventing disease progression and further episodes of AP. Although alcohol consumption and cholelithiasis are responsible in 60-80% of cases of AP(12, 13), the causes are multiple, and 10% of cases are classed as idiopathic, severely limiting definitive therapeutic options in this cohort of patients. This is in part because the pathophysiology remains poorly understood.

Metabolic profiling coupled with computational modeling has emerged as a tool for describing metabolic systems and is capable of characterizing different time-points in diseases(14, 15). This approach offers one potential avenue for improving the diagnosis and prognosis of AP in order to guide treatment and improve outcome. The determination of metabolite changes that describe a biological phenotype, based on either ^{1}H nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS), has been widely applied for global profiling to define diagnostic or prognostic biofluid profiles for physiological or pathological states(14, 16-18). However, to demonstrate that a metabonomic strategy for characterizing pathology has genuine translational capacity, it must be robust enough to cope with the complex clinical heterogeneity encountered in real world clinical environments such as the emergency room, which differ greatly from standardized and tightly controlled experimental environments. To our knowledge, this approach has yet to be prospectively applied to patients with AP presenting in the acute setting, although a small NMR-based comparison has been made between AP patients from an inpatient clinic (n=5) with healthy controls from an outpatient (n=5)(19). A key feature of AP is abdominal pain, a

common and often nonspecific presenting complaint for a large array of other surgical pathologies that require urgent treatment. However, 40% of all cases of acute abdominal pain are labeled, as being 'non-specific', and thus there is also an urgent need for early pathology specific biomarkers for causes of peritonitis. The aim of the current study was therefore to determine the diagnostic potential of a metabonomic approach in the diagnosis and prognostic staging of acute pancreatitis and to ascertain the clinical utility of this approach in the analysis of acute abdominal pain in the emergency setting.

MATERIALS AND METHODS

Study design. This was a prospective and observational control study of consecutive patients presenting with acute abdominal pain to a single center 'Accident and Emergency' unit at St. Mary's hospital, London, UK. All work was approved the by the ethical committee in St. Mary's Hospital, London (Rec 05/Q0403/201). Patients presenting with acute severe abdominal pain of less than 72 hours duration were enrolled. Patients who were pregnant, under the age of 18, who had undergone recent surgery or a blood transfusion were excluded. Patients with AP were identified by their medical history and examination, and an elevated amylase that was greater than three times the standard range (0-90 U/L). A pancreatic lipase assay was not routinely available for use within the hospital. Patients were investigated and managed as per the current British Society of Gastroenterology guidelines(20). Therefore, all patients underwent an abdominal ultrasound (USS) within 24 hours, and a Computed Tomography (CT) of the abdomen at 72 hours if no clinical improvement was obtained, or sooner if clinically indicated. Patients underwent clinical severity scoring using the modified Glasgow criteria(5) at 24 and 48 hours. All patients were given a diagnosis for their aetiology, and underwent an endoscopic retrograde cholangiopancreatography (ERCP) on the same admission if evidence of obstructive

biliary cholelithiasis was demonstrated. Patients with a previous history of pancreatitis (i.e. > 2 episodes) were subcategorized to have either recurrent or acute on chronic pancreatitis (CP). CP has a different mechanism and clinical presentation, and these patients were therefore excluded from the final analysis. The clinical diagnosis on discharge from hospital was used for patient phenotyping, as determined by the attending consultant physician. The aim was to achieve a practical representation of how clinical, biochemical and imaging data can be used to reach the diagnosis during the admission phase and to determine if a metabonomic method was able to augment this approach and to differentiate patients or subsets of patients according to presence, severity or cause of AP at an early stage. The 30 day follow up data were also recorded to confirm the final diagnosis. The control group was investigated according to the discretion of the attending surgeon. However, their diagnosis was confirmed on the basis of the clinical presentation, basic serum and urine biochemistry and on imaging by USS or CT or at the time of surgery with histopathological analysis. All patients had clinical observation data (pulse, temperature and blood pressure) and routine clinical biochemical data collected on admission and then at 24 hour intervals until the patient underwent surgery or was discharged. Samples were collected from each patient for the first five days of their admission. Samples for metabonomic analysis were collected as soon as possible after attendance of the patient in the emergency room or within 24 hours of admission to hospital. Patients not recruited prior to this cut off were excluded.

Sample collection and preparation.

Urine samples. Sampling was performed pre-catheterization wherever possible. If catheterized, a ‘clean catch’ sample was taken. Urine samples were kept on ice after collection prior to being stored frozen at -80 °C. Urine samples were thawed on ice, vortex and centrifuged,

they were prepared by adding 200 μL of phosphate buffer pH 7.4 containing 20 % of D_2O to 400 μL of urine. The mixture was transferred into a 5 mm outer diameter NMR tube. All ^1H NMR spectra were acquired using a Bruker DRX600 spectrometer (Rheinstetten, Germany) with a 5 mm TXI probe operating at 600.13 MHz. The field frequency was locked on to D_2O solvent. Primary acquisitions were made using a standard 1-D pulse program [Recycle delay (RD)-90^0-t_1-90^0-t_m-90^0- acquire free induction decay (FID)]. The 90° pulse length was adjusted to approximately 12 μs. A total of 64 scans were recorded into 32 K data points with a spectral width of 20 ppm. An exponential function was applied to the FID prior to the Fourier transformation, which resulted in a line broadening of 0.3 Hz.

Plasma samples. Plasma samples were collected in 5 ml aliquots and in heparinized lithium tubes and they were kept on ice before being centrifuged. The supernatant obtained was frozen at -20^0C. 400 μL of 0.9% saline solution were added to 200 μL of plasma. A 0.9% NaCl (wt/vol) solution was prepared with 10% of D_2O. ^1H NMR spectra of the plasma samples were acquired employing two 1D NMR experiments. These were a standard one-dimensional pulse sequence (as described for urine) and a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence [RD-90°-(τ-180°-τ)$_n$-acquire FID] giving a total spin-spin relaxation delay (2nτ) of 80 ms, determined by the number of spin echoes and τ= 400 μs. Typically in the standard one-dimensional and CPMG experiments, 256 and 128 transients were collected into 32 k data points respectively. All the method characteristics and details are detailed elsewhere(21).

Data Processing of NMR Spectra. All urine and plasma NMR spectra were automatically phased, baseline-corrected and referenced either to sodium 3-(trimethylsilyl) propionate-2,2,3,3-d4 (TSP; δ 0.0) or α-glucose (δ 5.23) for urine and plasma respectively using an in-house script: MetaSpectra for urine, and spec_preproc_v5 code for plasma(21). The spectra were imported to

Matlab and the region containing water and urea resonances (δ 4.7-6.2) and TSP (δ -0.1 – 0.1 for urinary spectra only) were removed. Peak alignment(22) and spectral normalization using a probabilistic quotient algorithm(23) were performed using an in-house Matlab script (The MathWorks, Natick, Massachusetts).

Multivariate Analysis. Principal component analysis (PCA) was applied to NMR spectral data sets to visualize outliers and data trends(24). Partial Least Square Discriminant Analysis (PLS-DA) was performed to optimize the classification of samples and aid identification of candidate biomarkers associated with clinical phenotype. Orthogonal-PLS-DA (OPLS-DA)(25) was also carried out to optimize recovery of potential biomarkers, i.e. compounds with high correlation and covariance with the class (e.g. disease severity or aetiology). Modeling was conducted in SIMCA-P v.12.1 using unit variance scaling. Pseudo-loadings plots were created with a color code projected onto the spectrum to indicate the correlation of the metabolites discriminating between the classes in each comparison, i.e. representing acute pancreatitis and other abdominal pain aetiologies. Red indicates high correlation and dark blue denotes no correlation with sample class. The direction and magnitude of the signals relate to the covariance of the metabolites with the class in the model. The quality of the models was assessed by the cumulative R^2 and Q^2, indicating the goodness of fit and the predictive power of each model using SIMCA software (Umetrics SIMCA-P+12.0.1 software). Metabolite identification was based on chemical shifts published in literature and Statistical Total Correlation Spectroscopy (STOCSY) of peaks(26) (e.g. STOCSY of urinary 3-hydroxyisovalerate is shown in Figure 1-IS). HMDB and COLMAR online databases were also used to confirm the chemical shift values, for all metabolites. Finally, where unknown metabolites were identified, or there was uncertainty in the metabolite identity, 2D NMR experiments, e.g. ^{1}H-^{1}H total correlation spectroscopy

(TOCSY) and ^{1}H -^{13}C heteronuclear multiple quantum coherence (HMQC) and heteronuclear single quantum coherence (HSQC), were performed for complete chemical structure elucidation. Parametric clinical data were analyzed by unpaired two-tailed Student's *t*-test using SPSS v. 19.0.

Exclusion of outliers and sample processing. A total of 46 patients were enrolled into the study (AP= 22, controls= 24). From this initial recruitment 5 patients had chronic pancreatitis (CP) and were excluded from the analysis as they represent a different pathology; an example of urinary CP spectrum is shown in Figure 2-SI. Additionally, strong outliers detected either by extremely unusual NMR spectra dominated by a few signals or by significant broadened signals due to e.g. excess proteinuria were removed. Finally, after removing the patients unable to provide samples the numbers of patients for urine analysis were: AP=13 and controls= 18, while for plasma; AP= 15 and controls= 21.

RESULTS

Clinical Data. Patients were aged 18 to 77 years old, with a median age of 42.5 years (SD=16.8). The remaining demographic data are provided in Table 1. The aetiologies of pain in the control group were heterogeneous reflecting the typical distribution of the causes of acute abdominal pain presenting to a typical accident and emergency department(27) (Table 1). Two patients in this cohort were not given a definitive diagnosis for their abdominal pain, and 3 patients had a primary biliary pathology but with a normal blood amylase level at presentation. The aetiologies of the AP group were either due to cholelithiasis (6/15) or alcohol (9/15). One of the cholelithiasis AP cohort had necrotic pancreatitis. The mean Glasgow score (a disease specific clinical risk score of severity ranging between 0 and 3, where 3 represents a severe

episode) was 2.07 at 24 hours and 1.07 at 48 hours indicating that this cohort of AP suffered moderately severe episodes. However, one patient with cholelithiasis AP and one patient with alcoholic AP diagnosis had a myocardial infarction during their admission, both of whom were treated with a standard acute coronary syndrome protocol. The APACHE II score(11) varied significantly between the groups, indicating the increased general severity of illness in the AP group compared to the control group. This was also reflected in the hospital length of stay (LOS) data, which differed significantly (average for controls = 3 and AP = 7; $p < 0.02$) between diagnostic groups (Table 1).

Characterizing the urinary phenotype of AP. The urinary ^{1}H NMR spectra from this clinical study demonstrated substantial inter-individual variation. Figure 1, shows urinary ^{1}H NMR spectra from an alcoholic AP patient (A) and a control patient (B). In general spectra from AP patients were characterized by high concentrations of urinary glucose, and ketone bodies. Unsurprisingly, large concentrations of acetate, ethanol, acetone, and ethyl glucuronide were also observed in this patient reflecting the underlying aetiology (Figure 1A). Ethanol was not consistently observed in the 6 patients with alcohol induced pancreatitis due to the varying volume of alcohol consumed between patients and time since ingestion (range 24 hours to 1 week). Furthermore, it was apparent from visual inspection across both groups (AP and controls) of urinary spectra that metabolites from common analgesic non-steroidal anti-inflammatory drugs (such as ibuprofen) were observed (data not shown), as might be expected in a cohort of patients presenting with abdominal pain. Signals from the drug acetaminophen metabolites were present in both groups, but they varied between patients; higher concentrations of acetaminophen metabolites can be observed in the cholecystitis sample in Figure 1B. NMR spectral profiles of acetaminophen and ibuprofen and their urinary metabolites have been published previously(28,

29). Urine specimens from several patients contained high concentrations of mannitol [δ 3.65(dd); δ 3.73(m); δ 3.77(d); δ 3.84(dd)], used as an excipient in the acetaminophen preparation and which was therefore correlated with acetaminophen and its related urinary metabolites [acetaminophen sulfate (AS)= δ: 2.18(s), δ 7.31(d), δ 7.46(d); acetaminophen glucuronide (AG)= δ 2.17(s), δ 3.62(m), δ 5.10(d), δ 7.13(d), δ 7.36(d); acetaminophen (A)= δ 2.16(s), δ 6.91(d), δ 7.25(d); N-acetylcysteine conjugate (NAC)= δ 1.86(s)]. The drug-related origin of the mannitol was supported by a Pearson correlation test(30) using all compounds in the profile and a selected driver peak for the mannitol signal (δ 3.805 ppm, (5-CH)) with an established cutoff of $p < 0.05$. The resulting graph showed mannitol signals and acetaminophen (A, AG, AS and NAC) were positively and highly correlated showing correlations values above 0.897 (from NAC 6.98(d)) with a p value of 0.039 (Figure 3-SI). This relationship was confirmed by ^{1}H NMR analysis of the intravenous acetaminophen preparation, a common analgesic given in the acute setting. Mannitol is a commonly used excipient in acetaminophen formulations(31). The regions containing mannitol (from δ 3.65 to δ 3.97) and acetaminophen metabolites (from δ 1.85 to δ 1.88; δ 2.14 to δ 2.2; δ 3.59 to δ 3.65; δ 7.11 to δ 7.18 and δ 7.29 to δ 7.49) were therefore removed from the raw data for the remaining analysis to minimize any confounding effects unrelated to the disease aetiology or severity.

Plasma NMR spectra. There were few obvious disease specific metabolites that varied systematically between raw spectra, however clear quantitative changes in the spectra were observed, particularly in concentrations of VLDL, LDL and non-esterified fatty acids. A comparison of 1D-NMR spectra from an AP and a control patient with appendicitis is shown in Figure 2 illustrating the gross differences in the lipoproteins. Likewise, a comparison of CPMG

raw spectra from different patients; an AP and control patient diagnosed with diverticulitis, is presented in Figure 4-SI where differences of VLDL were observed.

Multivariate modeling of patients with severe abdominal pain. Three main analyses were performed based on clinical criteria; 1) diagnosis of AP (AP vs. all controls); 2) aetiology (alcohol consumption vs. cholelithiasis); and 3) AP severity, based on a validated modified Glasgow score comparison of mild (Glasgow score 1) versus moderate to severe (Glasgow score 2 and 3). It was not possible to build predictive models for organ failure or mortality based on the small numbers of patients. PCA, PLS-DA and O-PLS DA models were therefore created to determine the capacity of the metabolic profiles to classify the sample set according to the above outcomes and to determine how well the model could predict the class of samples according to the principal diagnosis. PCA models did not show clear separation for any of the clinical scenarios in this study, due to the extreme heterogeneity of the data. PLS-DA plots for each clinical scenario in urine and plasma (CPMG- and 1D-spectral) data (Figure 3) showed valid predictive models for all, with relatively robust Q^2 values that were supported by permutation tests, for each of the biofluids with the exception of prediction of aetiology in the plasma (1D-plasma and CPMG) data set. Discriminatory metabolites were found in the Orthogonal-PLS-DA pseudo-loadings plots for each of the three sets of models for AP prediction, severity and aetiology in urine and plasma (Figure 4) and differed in composition between models assessing different outcomes. Discriminant metabolites found in plasma and urine for the three clinical scenarios are summarized in Table 2 and discussed below.

1. Diagnosis. Consistent changes in concentrations of urinary and plasma metabolites were found between the AP and the control group. Plasma acetone, D-3-hydroxybutyrate, acetoacetate, glucose, lipid signals (CH3CH2-, CH2CH=CH) fraction and plasma choline

concentrations were increased in AP patients. Low urinary levels of hippurate, creatine and plasma valine and alanine were also characteristic of the AP group. Since hippurate is a well-established gut microbial co-metabolite derived from glycine conjugation of benzoic acid(32, 33), a sub-group analysis of antibiotic treated patients (n=10; 5controls/5AP) and non-antibiotic (n=21; 12controls/9AP) treated patients was performed to determine if antibiotic ingestion in the AP diagnostic model was responsible for the difference in urinary excretion of hippurate between the two groups (figure 5/ table1-SI). The PLS-DA analysis demonstrated class separation according to antiobiotic usage (R^2= 0.85, Q^2= 0.47) based on modulation of a combination of metabolites including decreased urinary citrate, methylamine and creatinine suggesting that antibiotic use did have a systemic metabolic impact in this group of patients. Some of the differentiating signals derived directly from antibiotics (e.g. metronidazole(34)). Concentrations of 3-hydroxyisovaleric acid previously reported to be a gut microbial co-metabolite(35) were decreased in the antibiotic group. However, hippurate concentrations did not vary significantly (driver peak= δ^1H 7.85 (6-CH); correlation (r)= -0.29; p= 0.11) between groups suggesting that the AP pathophysiology has a greater effect on this mammalian-host co-metabolite than antibiotic use. In addition to signals from analgesics and NSAID metabolites (Figure 5-SI), the multivariate urinary model of antibiotic usage (Figure 5/ Table 1-SI) shows unknown compounds whose excretion was increased in the antibiotic group, which are most likely to be xenometabolites. Antibiotic usage was associated with decreased urinary concentrations of 3-hydroxyisovalerate, methylamine, citrate and creatinine.

In order to compare the robustness of the metabonomic models in urine and plasma, ROC curves were created (Figure 6A-B). The urine ^{1}H NMR model (AUC = 0.91) produced a stronger diagnostic model than plasma (AUC = 0.86). The urinary multivariate model had a high

sensitivity with True Positive Rates (TPR) of 0.5-0.6, low False Positive Rates (FPR) of < 0.2 and a classification rate of ≥80%. Additionally 10 metabolites (5 from each biofluid) were combined in order to perform a model and the respective ROC curve from both (Figure 6C). Interesting the curve showed better results in the prediction of AP diagnosis were (AUC= 0.96) compared to the global profiles of each biofluid separately. Moreover the OPLS-DA cross validated showed better quality parameters (Figure 6D, R^2=0.76, Q^2=0.59). Finally serum amylase data had a sensitivity of 100% and a specificity of 95%, which although comparable to the metabonomic data have little clinical significance for this study as it was used as a diagnostic criterion. Moreover, the use of serum amylase as a diagnostic does not contribute new information towards understanding the aetiology of AP or augment patient stratification with respect to treatment. The metabonomic data therefore provide an improved and comprehensive systems level overview of an individual's metabolic response to pancreatic injury than a single biomarker.

2. Aetiology: Multivariate models of alcohol versus cholelithiasis AP had a good classification capacity (R^2Y from 86 to 96%) but variable predictive capacity; Q^2 from no predictive capacity to 41% (Figure 3, ii). The urinary ketone body excretion of acetone, acetatoacetate and urinary 3-hydroxyisovalerate was increased in the alcoholic AP patients compared to AP patients with cholelithiasis. The models generated from the plasma (CPMG and 1D) spectra carried a negative predictive value in the first component, indicating that there was no clear diagnostic signature of aetiology based on the low molecular weight components of plasma and thus only worthy for explaining classification of the dataset samples. However, OPLS–DA from urinary ^{1}H NMR spectra generated a stronger model (R^2Y = 0.76, Q^2 = 0.44), confirming that acetone and acetoacetate are of importance for the diagnosis of alcoholic AP (Figure 4c, i). Nevertheless

glucose which was highly correlated with cholelithiasis and D-3-hydroxybutyrate increased in the alcohol aetiology were only representative for the sample classification (Figure 4b-c, ii).

3. Severity of AP: Multivariate models for disease severity were created using the Glasgow score as a surrogate marker. Both plasma and urine produced robust predictive models, but plasma CPMG data generated the strongest model ($R^2Y = 0.96$, $Q^2 = 0.64$), with tight clustering of those AP patients with mild disease, indicating that it was the low molecular weight components of plasma that were modulated most according to disease severity. Plasma metabolites such as creatinine, alanine, valine, choline and N-acetyl signals from α1-acid glycoprotein were inversely related to disease severity, while increased plasma glucose strongly correlated with severity (Figure 4, iii). Of note, in patients with more severe disease, it was also possible to detect biomarkers of plasma expanders (such as the colloid 'volplex') which are commonly used to resuscitate patients with hypovolaemic or septic shock. However, removal of these treatment-related resonances from the models did not impact on the model quality or endogenous correlates with disease severity.

Finally, for control group data analysis, patients were divided in two principal cohorts: A) those with cholelithiasis and B) colonic inflammation i.e. patients with diverticulitis, appendicitis, Crohn's disease and intestinal ischaemia. Patients that did not fit the defined criteria for either of cholelithiasis or colonic inflammation group were excluded for the comparison (e.g. patients with urosepsis or non-specific abdominal pain without a diagnosis). Robust PLS-DA models for urine and plasma indicated that the cholelithiasis sub-group was metabolically distinct from the colonic inflammation sub-group (Figure 7). Urinary concentrations of metabolites such as malonate, and 3-hydroxyisovalerate, identified from the OPLS-DA loadings plots, were higher in the group with colonic inflammation while plasma concentrations of valine,

leucine, choline, lactate, and lipids were higher in the cholelithiasis group along with urinary 3-indole acetate. These discriminant metabolites are summarized in Table 3.

DISCUSSION

The complex aetiology of acute pancreatitis and the exact mechanism of the consequent systemic inflammatory response (SIRS) up-regulation have yet to be fully elucidated. This is compounded by the fact that the main symptom of AP is acute abdominal pain, which is a common, non-specific and a highly subjective presenting complaint in the emergency setting. There are a lack of non-invasive biomarkers for the diagnosis of the underlying cause of peritonitis, and as a result up to 40% of cases of acute abdominal pain are still labeled as 'non-specific'(27). Since there is no effective treatment for AP other than supportive therapy, enhanced mechanistic knowledge of the disease is critical for guiding more effective therapeutic strategies.

Diagnosis of AP. Here we demonstrated that multivariate analysis of both urinary and plasma ^{1}H NMR spectra was able to accurately stratify patients presenting with acute abdominal pain in a clinical setting into those with AP and those with other heterogeneous causes of abdominal pain (Figure 3, i). Urine provided a stronger predictor of AP diagnosis (ROC curve data, AUC= 0.91, Figure 6A) than plasma (ROC curve data AUC= 0.87, Figure 6B). The top 5 signals discriminating the AP and control group in urine and in plasma were combined and this selected panel outperformed the models for either urine or plasma individually, built from the global profile (ROC curve data AUC= 0.96, Figure 6C) observed even for the cross validated OPLS-DA plot.

AP is associated with a disruption in glucose metabolism caused by the severe stress response, SIRS and the resulting insulin resistance and altered pancreatic function. Insulin resistance

activates secondary metabolic pathways with accelerated adipose tissue triglyceride hydrolysis causing increased ketogenesis and ketoacids visible in blood and urine. Severe vomiting, anorexia, dehydration and alcohol use can also cause high levels of ketone bodies (acetone, D-3-hydroxybutyrate and acetoacetic acid), which are excreted in urine. This is supported by previous work which has found that fasting blood glucose and insulin levels are higher in AP patients than in controls(36). A striking observation was the presence of non-esterified fatty acids (NEFA) in the plasma of patients with AP or severe illness as governed by the APACHE II score (Figure 2). Changes in the lipid profile in our study (high levels of low-density lipoprotein [LDL; 0.86, 1.25(m)], and unsaturated lipid: CH2CH=CH [1.97(m)]) are consistent with known disruptions in lipid metabolism in AP, and indeed hypertriglyceridaemia is a rare, but well established, cause of AP with a much more malignant phenotype and a poorer outcome(37, 38). During the acute phase response, plasma triglyceride levels rise because of increased VLDL secretion as a result of adipose tissue lipolysis, increased de novo hepatic fatty acid synthesis and suppression of fatty acid oxidation. With more severe infection, VLDL clearance decreases secondary to decreased lipoprotein lipase and apolipoprotein E in VLDL(39). Furthermore, rodent models using a lipopolysaccharide endotoxin challenge have demonstrated that this induces an increase in plasma NEFA and large, triglyceride-rich VLDL(40). Thus it is very likely that the striking rise in plasma VLDLs seen here relates to the generalized severity of the systemic response rather than AP specific changes. Hyperlipidemia may also cause an elevation of plasma choline associated with triglyceride breakdown, or more likely phosphatidylcholine, which in turn leads to formation of fatty acids and an exacerbation of the ketonuria. In general these changes are commonly observed in diabetes(41).

To our knowledge lower plasma concentrations of alanine and valine in the AP group than in the controls, observed in the current study have not been reported previously for AP, although decreased concentrations of amino acids (leucine, isoleucine, valine and alanine) have been found in both rodent models of chronic pancreatitis (CP)(42) and in humans(43) and lower levels of valine have been found in CP patients when compared to acute necrotizing pancreatitis.(44) In contrast, other studies have reported elevated alanine and creatine levels in CP patients compared to healthy subjects(45). It is perhaps more relevant that both alanine and valine are components of pancreatic juice(46), and their fall in circulating concentrations may be related to an exocrine pancreatic insufficiency that could be characteristic of AP compared to other intestinal or biliary conditions presenting with abdominal pain.

A significant anti-correlation was found for hippurate and AP, which was only partially accounted for by antibiotic usage (Figure 5/ Table 1-SI). Since hippurate, a well-established gut microbial co-metabolite and biomarker of hepatic reserve(33), was not significantly associated with antibiotic usage, this supports the theory that SIRS in AP leads to mesenteric hypoperfusion(47) which in turn may disrupt the gut luminal environment and subsequently gut microbial function. This also fits with the theory of gut bacterial translocation in AP, which is thought to explain the high rates of enteric bacteria identified in cases of pancreatic necrosis and sepsis. However, we are unable to exclude the possibility that altered hepatic glycine conjugation of Benzoyl CoA may also explain this finding. These data were also modeled without inclusion of drug metabolites, and there is a possibility that they will exert pharmacological effects on endogenous metabolic pathways.

Aetiology of AP caused by alcohol consumption and cholelithiasis. The two causes of AP in this series were alcohol (9/15) and cholelithiasis, which is representative of the epidemiology of the condition. Metabolites such as acetate, acetone, acetoacetate and 3-hydroxyisovalerate were predictive for alcohol induced AP. It is well established that alcohol consumption influences the regulation of key pathways such as gluconeogenesis. Ketone bodies are products of the oxidative pathway of alcohol metabolism(48) and are thus concordant with this observation discriminating the alcohol-induced AP patients, although since there were no dietary records for the patients, the possibility that anorexia, associated with severe abdominal pain, may have been systematically different between the groups and therefore a contributor to the ketone body profile.

AP severity based on the Glasgow severity score. As is typical, there was a range of severities, with a significant difference in the hospital length of stay (LOS) and APACHE scores between the control and AP groups (Table 1). This was also reflected in the both in routine biochemistry, and hematological parameters measured between groups (table 2-SI) and in the raw NMR spectra, which were highly variable in structure (Figures 1 and 2). These data provide a strong metabolic phenotype of the severity of illness, and a clear metabolic description of AP disease severity. AP patients with moderate to severe AP disease (Glasgow 2-3) had higher levels of glucose and reduced plasma relative concentrations of alanine, valine, choline, the acetyl signal from α1-acid glycoprotein (NAC1), and urinary creatinine. As previously described, this combined cohort of metabolic changes is suggestive of insulin resistance(49) and probable failure of pancreatic function, with the NAC1 most likely reflecting generalized inflammation(50).

Differentiation of metabolic phenotypes within the control group. It was possible to distinguish between subclasses of the control group (Figure 7 / Table 3) using the metabolic profiles, which is of importance to the clinical setting. Patients with cholelithiasis demonstrated higher concentrations of lactate, choline, valine, leucine, and lipids in the plasma, and increased urinary indoleacetate levels. Hypertriglyceridaemia, often found in overweight and insulin resistant individuals is associated both with an increased risk of cholelithiasis disease(51) and malignant AP. In addition increased plasma lactate levels have been shown to be associated with malignant diseases of the liver and biliary tract(52). Indoleacetate is a metabolite produced from tryptophan by several intestinal bacteria. High levels of indoleacetate have typically been related to gastro-intestinal cancer and hepato-biliary tract cancer; and high concentrations have been observed in patients with cirrhosis, diabetes and cholelithiasis occasionally(53). Independent studies suggest that low plasma choline concentrations are associated with cholelithiasis and colonic inflammation(54, 55). Interestingly, our data demonstrate that when these two groups are compared, higher concentrations of choline are observed in the cholelithiasis group. Moreover, in our study a negative correlation of branched-chain amino acids (valine and leucine) with the colonic inflammation group was observed. This is in keeping with other data suggesting that leucine and valine were found in lower concentrations in the colonic mucosa of patients with inflammatory bowel disease(55).

Increased amounts of the urinary metabolites malonate, and 3-hydroxyisovalerate were observed in our study in the group with colonic inflammation. Malonate is a carbon source utilized by a variety of intestinal bacteria(56). Interestingly common gut microbial activity metabolites such as hippurate or phenylacetylglutamine (PAG) were not found to be of diagnostic importance in patients with colonic pathology.

Although this study has several limitations, which must be accounted when interpreting the data, the differential metabolic phenotype for AP is convincing as all the models presented matching results and were supported by the overall permutation test. Of note, however, the numbers of patients was relatively small and the study was not powered for mortality. Larger numbers of patients with greater severity of illness are required to build more accurate prognostic models that are not based on surrogate markers of disease severity and to explore the power of global metabolic profiling to distinguish between rarer causes of AP (e.g. viral, drug induced or hypertriglyceridaemia). Also, males were statistically over represented in the AP group, although this is a reflection of the epidemiology of the disease. Nevertheless, despite these limitations, we have demonstrated that metabolic profiling has considerable clinical utility in the management of acutely unwell patients. Despite the potential of multiple confounding signals deriving from drugs or excipients such as mannitol and antibiotics or from intravenous fluids given to patients during their resuscitation (such as volplex), all of which can bias the multivariate models, we have shown clear metabolic differentiation based on presence, severity and aetiology of AP. The diversity of signals form therapeutics and other confounding sources could however reflect a great strength of this approach however, as the metabolic systems medicine is able to objectively account for the impact of therapeutic interventions. ^{1}H NMR spectroscopy is a robust and highly reproducible high throughput technique for untargeted analysis of complex pathologies such as AP and SIRS. However, we also perceive that the use of other throughput analytical platforms such as liquid chromatography coupled to mass spectrometry (LC-MS) would contribute to increasing the sensitivity and specificity of diagnosis. It is likely that much of the diagnostic power of this model relates to the significant severity of illness encountered in patients when compared to generic peritonitic patients, as found in the variation between the APACHE II

scores and the hospital LOS (Table 1). The lipid signature appears to be extremely important in making this differentiation and the overall result of this scoping study suggests that a global metabolic profiling approach has significant potential for the early detection and diagnosis of critically unwell patients.

CONCLUSIONS

This analysis deployed a metabolic profiling strategy to determine a diagnostic and prognostic phenotype of AP reflective of the systemic injury, based on an untargeted combinatorial panel of metabolic biomarkers for acute pancreatitis in the emergency setting. Moreover, it is able to metabolically phenotype patients with severe disease and to measure the influence of drug treatments and therapeutic interventions on the pathological phenotype. In particular, we found the plasma lipid signature to be important in the clinical diagnosis of AP alongside valine, alanine and hippurate, differentiating AP from other common causes of abdominal pain. Information about drug metabolism, dietary influences on disease biology and the possible contribution of the gut microbiota in the disease mechanism was also obtained using a metabonomic framework.

Finally, this work has shown clinical utility of metabolic phenotyping approaches in the diagnosis of acute abdominal pain and in the critical care setting. Further work is required in much larger patient cohorts to further validate these findings and to refine the metabolic signature attributable to underlying aetiologies.

FIGURE LEGENDS.

Figure 1. Typical 600 MHz ^{1}H NMR spectra of urine obtained from a patient with acute pancreatitis and alcoholic background (A) and a control patient with cholecystitis (A). Sections

from the NMR spectra (raw data) show the metabolite differences describing the physical condition. Key: 1. Hippurate; 2. Acetaminophen sulphate (AS); 3. Acetaminophen glucuronide (AG); 4. Acetaminophen (A); 5. N-Acetylcysteine conjugate of acetaminophen (NAC); 6. Creatinine; 7. Guanidinoacetate; 8. Glycine; 9. Trimethylamine-N-oxide (TMAO); 10.Dimethylamine; 11. Citrate; 12. D-3-hydroxybutyrate; 13. Ethanol; 14. Acetone; 15; Alanine; 16. Succinate; 17. 3-hydroxyisovalerate; 18. Malonic acid; 19. Creatine.

Figure 2. 600 MHz ^{1}H NMR spectra (δ -0.0 to 4.7) of plasma from: (A) Pancreatitis patient and (B) Control patient diagnosed with appendicitis. Abbreviations: LDL1 and VLDL1 refer to the terminal CH_3 groups of fatty acids in low-density and very-low-density lipoproteins, respectively; HDL refers to the C18 signal from cholesterol in high-density lipoprotein; NAC1 and NAC2 refer to composite acetyl signals from α1-acid glycoprotein.

Figure 3. Scores plot of the first versus second component for Partial Least Square Discriminant Analysis for A) urine, B) Plasma – CPMG and C) 1D Plasma for three different comparisons: i) AP vs control (diagnosis), ii) aetiology of AP patients; alcoholic vs cholelithiasis background, iii) severity of AP group based on Glasgow score; Mild AP severity (Glasgow 1) vs Moderate to severe AP (Glasgow 2-3). Key: *, means the value for the first component only, unit variance scaling; two principal components; R^2 and Q^2 values indicate variance explained and predictivity of the model respectively; (i) ● AP samples (An=13, B-Cn=15); ■ Control samples (An=18, B-Cn=21). (ii) ● Alcoholic AP samples (An=8, B-Cn=10); ▲ Cholelithiasis AP samples (An=5, B-Cn=6). (iii) ■ Glasgow 1 group (An=8, B-Cn=9); ▲ Glasgow scores 2 and 3 group (An=5, B-Cn=7).

Figure 4. OPLS-DA coefficient plots, peak signals related to the discrimination between i) diagnostic classes: signals indicating higher concentration in AP group (upwards) vs controls (downwards); ii) comparison of the main causes of pancreatitis disease: cholelithiasis (signals oriented upwards) vs alcohol consumption (downwards) and iii) severity of the disease based on Glasgow score: Glasgow score 1 (upward signals) vs Glasgow score 2 and 3 (downward). The color indicates significance of the signal association with class, the hotter the color the higher the correlation with class. All comparisons for A) urine, B) plasma-CPMG and C) 1D-plasma used one correlated component, 1 orthogonal component in matrix X and 0 in class Y. Key: *, means the value for the first component only, unit variance scaling; two principal components.

Figure 5. OPLS-DA coefficients plot; peak signals related to the discrimination between antibiotic consumers (top) group vs non-antibiotic takers (bottom). Models were created using unit variance scaling, R^2=0.85, and Q^2= 0.47; ● Antibiotic takers (n=10 [5controls/5AP]); ■ non-antibiotic samples (n=21 [12controls/9AP]).

Figure 6. ROC plot for A) Urinary dataset (n=31 [18controls / 13AP]); B) Plasma-CPMG (n=36 [21controls / 15AP]); C) Combination of 10 metabolites in total from urine (guanine at 7.67(s), hippurate at 7.54(t), 7.64(t), 7.84(d), creatine at 3.02(s) and unknown metabolites at 2.73(s), and 4.55(d), and CPMG plasma (valine at 1.04(d), alanine at 1.47 (d), and lipid compounds at 0.91(m), 1.25(m) and 1.97(m)); OPLS-DA cross-validated from the previous ROC curve (R^2=0.76, Q^2=0.59, n=29 [16controls / 13AP]).

Figure 7. PLS-DA scores plot for A) urine, B) Plasma – CPMG and C) 1D Plasma for patients comparing cholelithiasis vs colonic inflammation (including diverticulitis, appendicitis, Crohn's disease and ischaemia), within the control group (Table 1), B) OPLS-DA loading plot, for the

discrimination between patients with cholelithiasis (upward facing signals) vs. patients with colonic inflammation (downward). The model was built using unit variance scaling. R^2 and Q^2 values signify variance explained and predictivity of the model respectively; ▲ cholelithiasis patients (A-B-Cn=5); ● patients with colonic inflammation (A-B-Cn=9).

Figure 1.

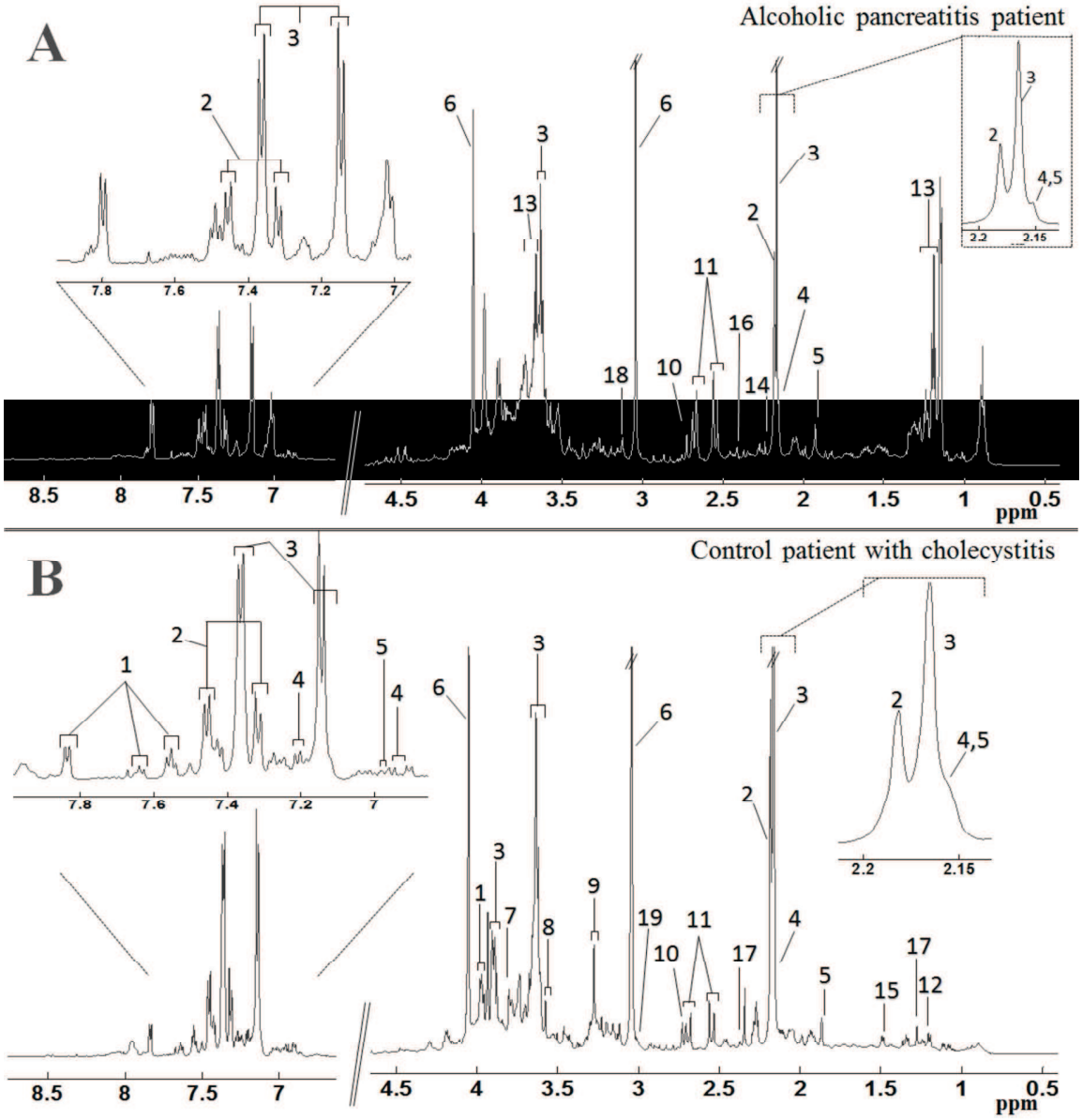

A
Alcoholic pancreatitis patient
B
Control patient with cholecystitis
ppm

Figure 2.

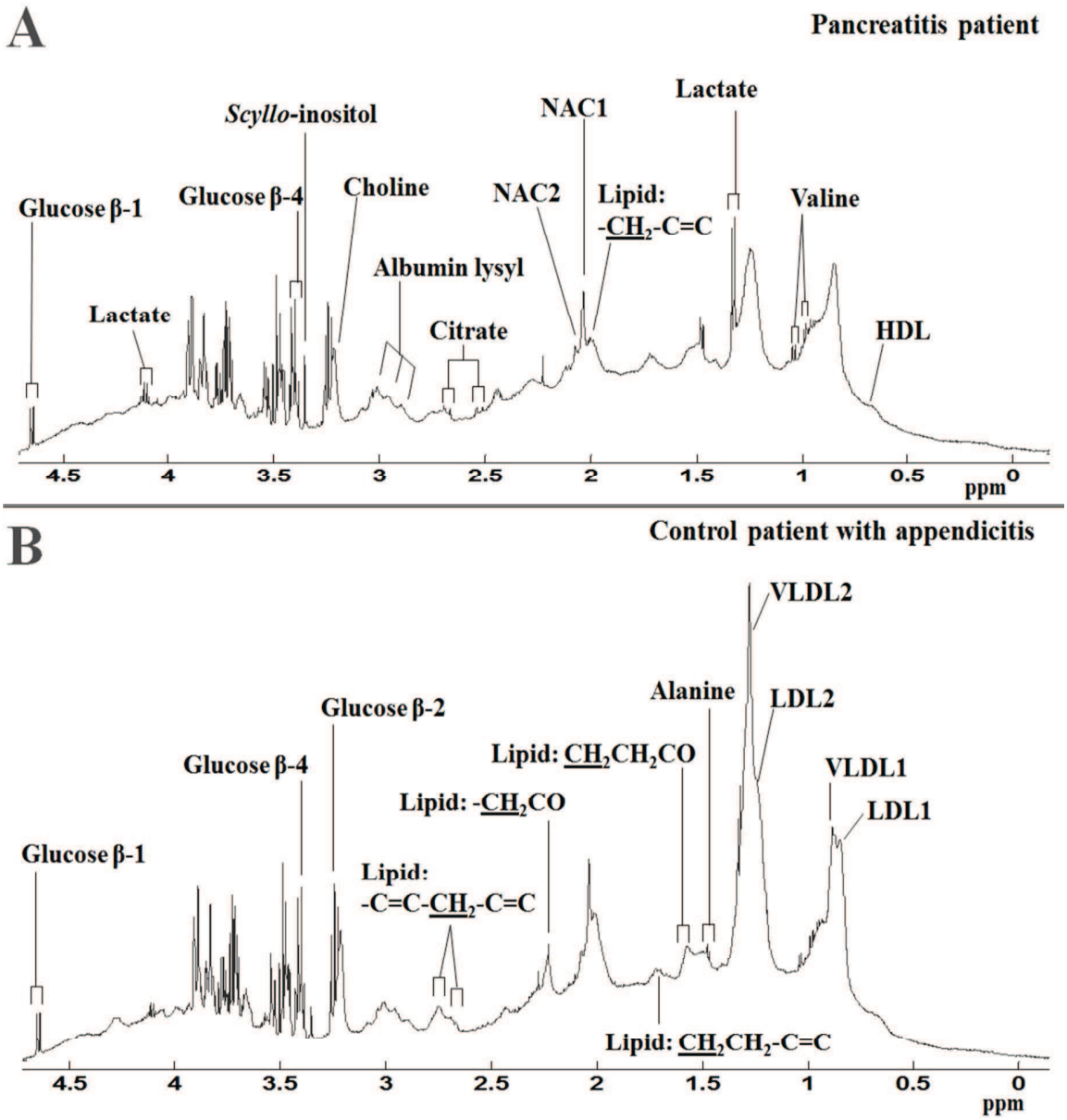
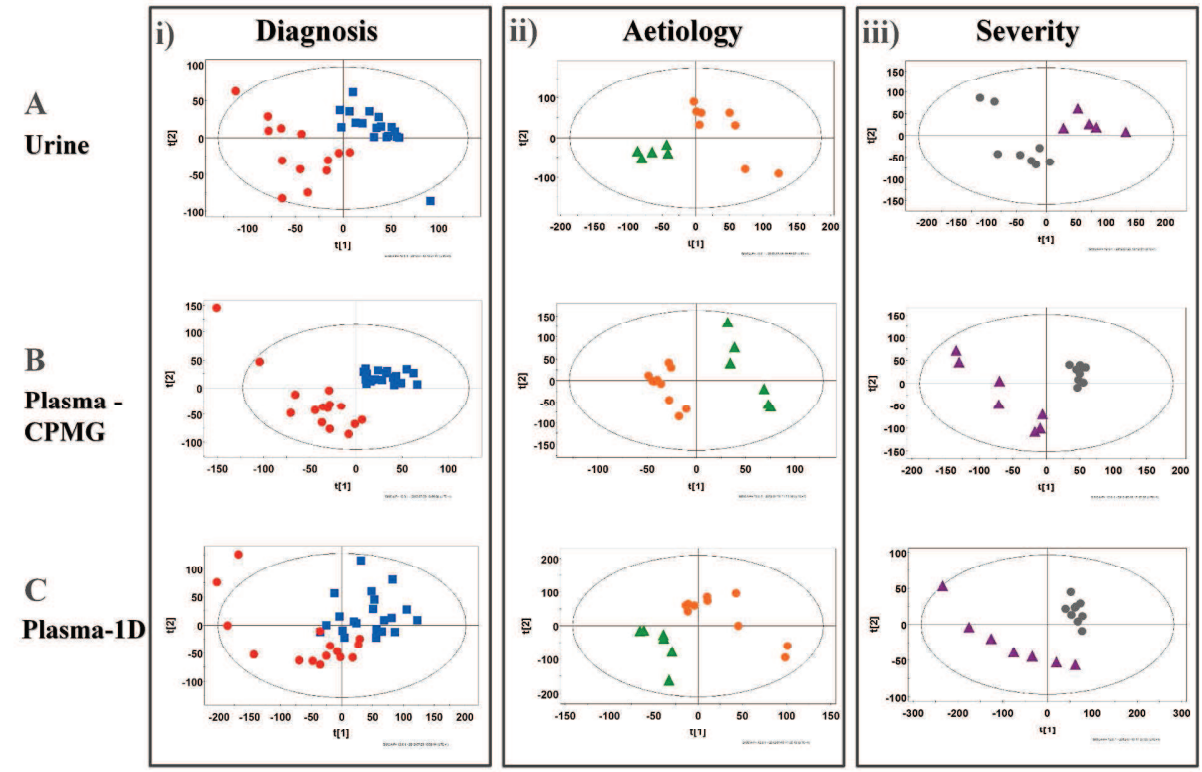
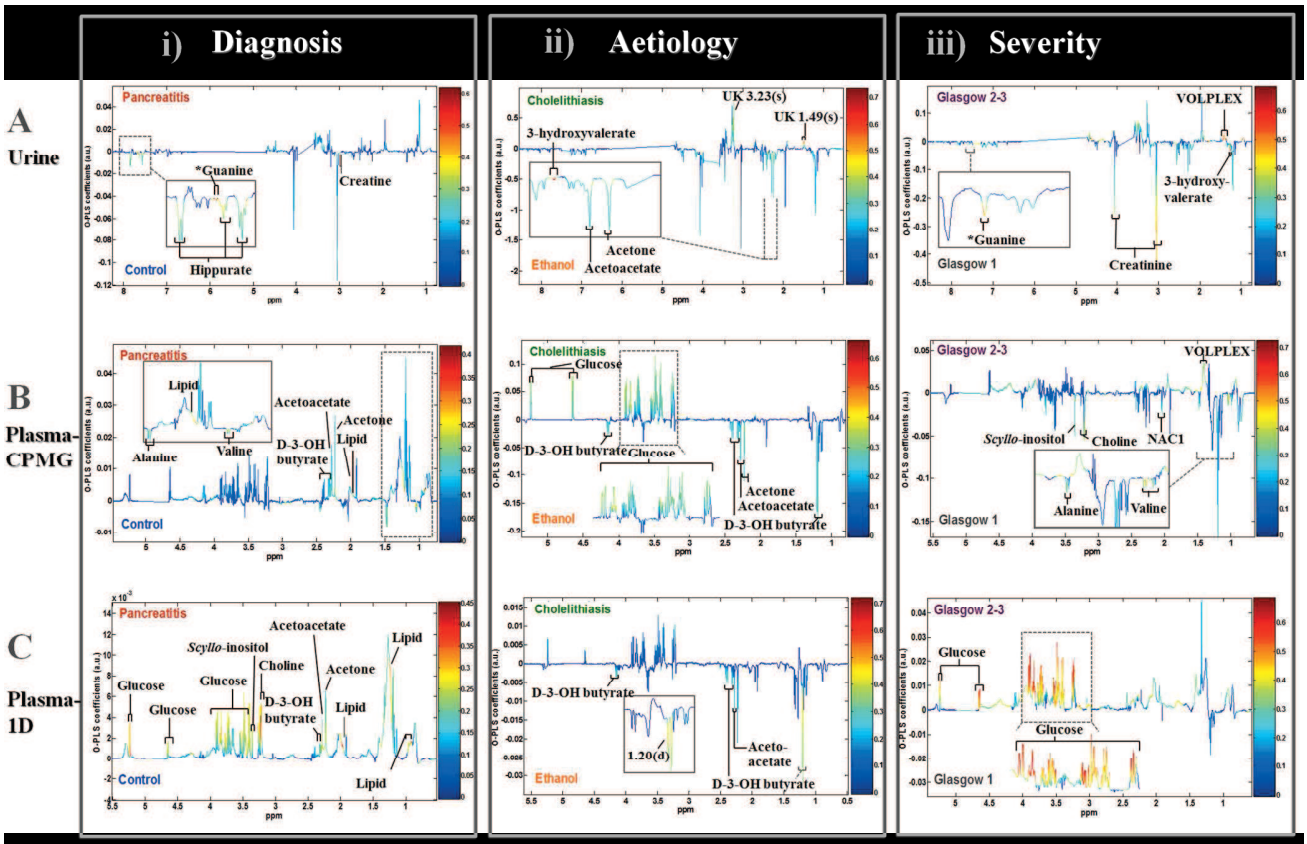

A
Pancreatitis patient
Scyllo-inositol
NAC1
Lactate
Glucose β-1
Glucose β-4
Choline
NAC2
Lipid:
-CH2-C=C
Valine
Albumin lysyl
Lactate
Citrate
HDL
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
ppm
B
Control patient with appendicitis
VLDL2
Glucose β-2
Alanine
LDL2
Glucose β-4
Lipid: CH2CH2CO
VLDL1
Lipid: -CH2CO
LDL1
Glucose β-1
Lipid:
-C=C-CH2-C=C
Lipid: CH2CH2-C=C
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
ppm

Figure 3.



	Diagnosis		Aetiology		Severity	
	R^2	Q^2	R^2	Q^2	R^2	Q^2
Urine	0.84	0.35	0.90	0.24	0.87	0.36
CPMG	0.89	0.55	0.88*	-0.10*	0.96	0.64
1D-Plasma	0.59	0.26	0.51*	-0.10*	0.92	0.33

Figure 4.



	Diagnosis		Aetiology		Severity	
	R^2	Q^2	R^2	Q^2	R^2	Q^2
Urine	0.84	0.48	0.76	0.47	0.88	0.43
CPMG	0.89	0.34	0.89*	-0.03*	0.96	0.26
1D-Plasma	0.60	0.19	0.51*	-0.33*	0.92	0.18

Figure 5.

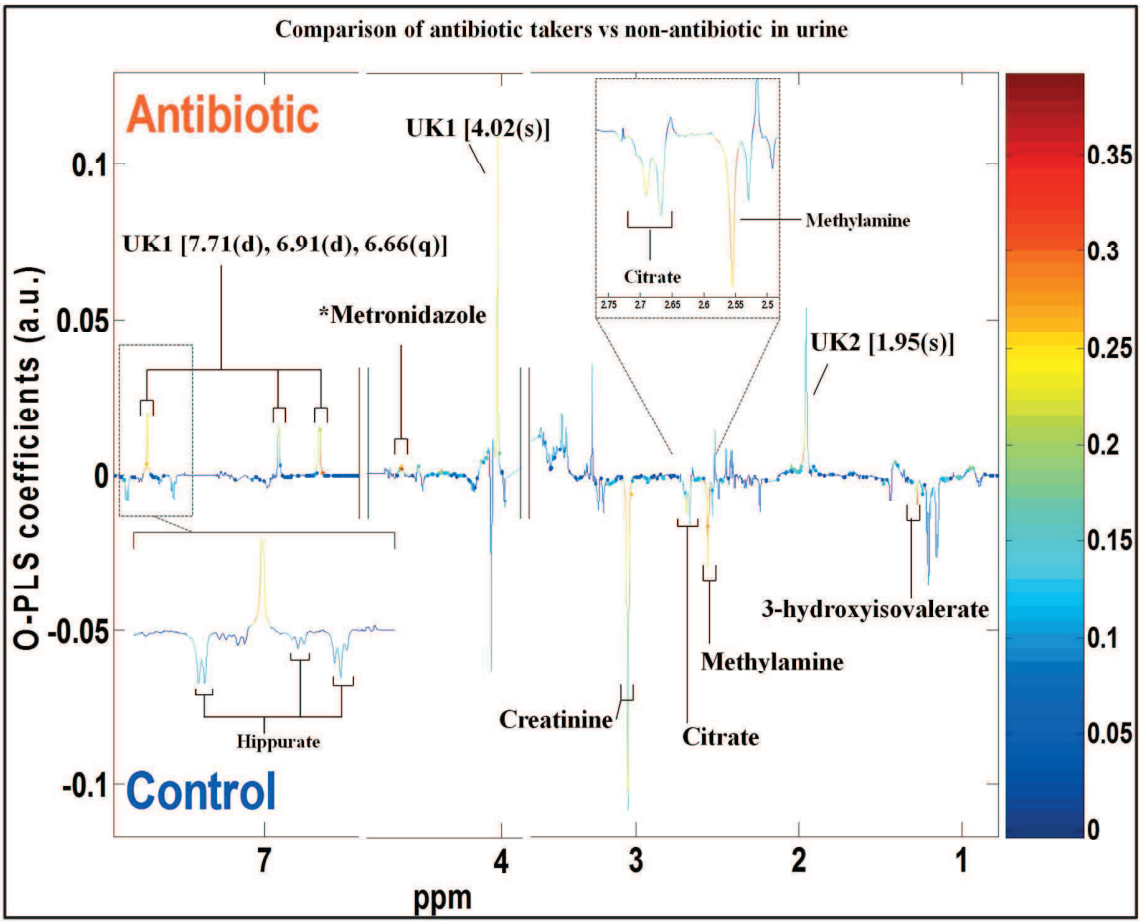

Comparison of antibiotic takers vs non-antibiotic in urine
Antibiotic
Control
UK1 [4.02(s)]
UK1 [7.71(d), 6.91(d), 6.66(q)]
*Metronidazole
UK2 [1.95(s)]
3-hydroxyisovalerate
Methylamine
Creatinine
Citrate
Hippurate
Methylamine
Citrate
O-PLS coefficients (a.u.)
0.1
0.05
0
-0.05
-0.1
7
4
3
2
1
ppm
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0

Figure 6.

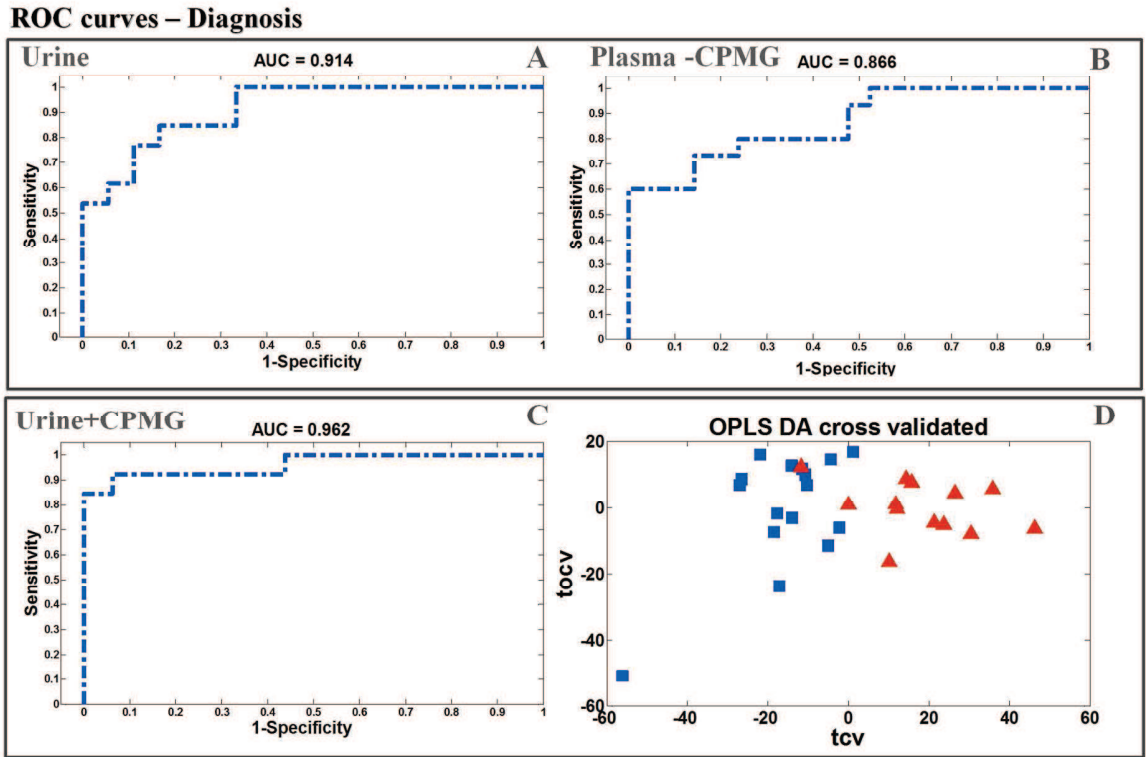
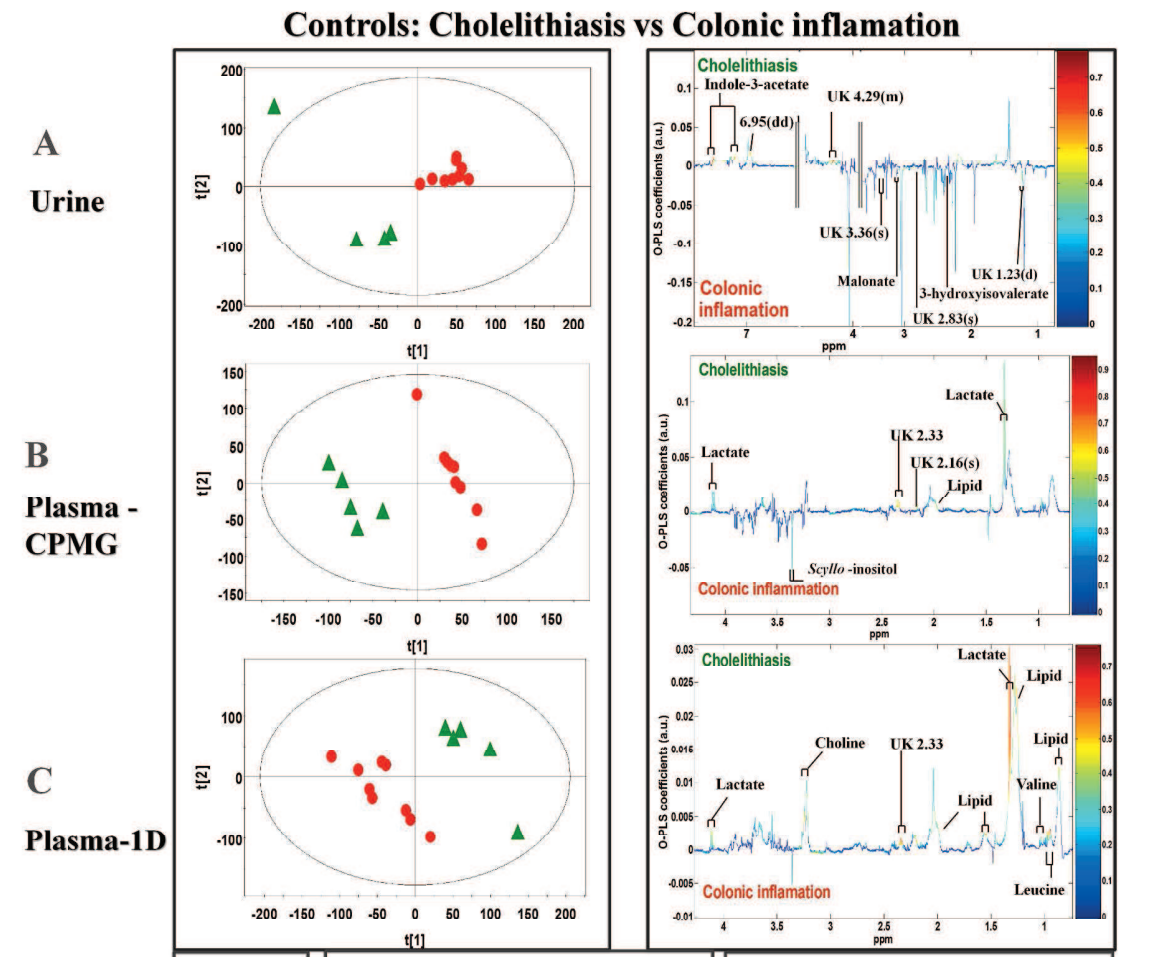

ROC curves – Diagnosis
Urine
AUC = 0.914
A
Plasma -CPMG
AUC = 0.866
B
Urine+CPMG
AUC = 0.962
C
OPLS DA cross validated
D
Sensitivity
1-Specificity
tocv
tcv

Figure 7.



	PLS-DA		OPLS-DA	
	R^2	Q^2	R^2	Q^2
Urine	0.91	0.60	0.91	0.50
CPMG	0.98	0.24	0.98	0.31
1D-Plasma	0.94	0.29	0.93	0.19

TABLES.

Table 1. Summary of clinical data.

		Control (n=23)	Pancreatitis (n=15)	*p value*
Age (Median, range)		42 (18-77)	50 (22-70)	0.980
Sex	F:M	12:11	3:12	0.001
	Cholecystitis	1		-
	Biliary colic	1		-
	Ascending cholangitis	1		-
	Appendicitis	5		-
	Diverticulitis	3		-
	Colitis	2		-
	Adhesions	1		-
	Unknown	2		-
	UTI	3		-
	Renal colic	1		-
	Caecal volvulus	1		-
	Gastritis	1		-
	Mumps orchitis	1		-
	Gall stone pancreatitis		6	-
	EtOH pancreatitis		9	-
Amylase (Mean, SD)		62.1 (73.7)	874.6 (841.8)	1.3E-06
Symptom duration (Mean, SD)		33.4 (24.1)	41.9 (23.8)	0.860
Glasgow 24 hrs (Mean, SD)		0	2.1 (1.8)	-
Glasgow 48 hrs		0	1.1 (1.3)	-
Apache 24 hrs (Mean, SD)		3.3 (3.2)	9.6 (6.3)	0.014

Apache 48 hrs	1.9 (3.0)	5.8 (3.9)	0.140
LOS (Median, range)	3 (0-14)	7 (2-132)	0.020

Data presented is the average (range or SD). Data were analyzed by unpaired two-tailed Student's t-test.

Table 2. Statistically significant metabolites and their corresponding chemical shifts, in the three main comparisons; **1)** AP group vs. controls (diagnosis), according to the **2)** AP aetiology cholelithiasis vs. Ethanol and **3)** AP severity mild AP (Glasgow 1) vs. moderate to severe AP (Glasgow 2-3).

Metabolite	**Chemical shifts (ppm) with multiplicities**	**AP group vs. controls**	**Matrix**	**driver (δ)**	**Correl. (*r*)**	***p* value**
Hippurate	3.96(d); 7.54(t); 7.64(t); 7.84(d)	↓	Urine	7.64	-0.584	0.001
Creatine	3.02(s); 3.92(s)	↓	Urine	3.02	-0.508	0.004
Valine	0.98(d); 1.04(d); 2.26(m); 3.60(d)	↓	CPMG	1.03	-0.356	0.033
Alanine	1.47(d); 3.77(q)	↓	CPMG	1.47	-0.411	0.013
Lipid CH_3CH_2	0.93(m)	↑	CPMG 1D-Plasma	0.91	0.619 0.611	0.000 0.000
Lipid (mainly LDL)	1.25 (m)	↑	CPMG 1D-Plasma	1.245	0.420 0.540	0.011 0.001
Lipid $CH_2CH{=}CH$	1.97(m)	↑	CPMG 1D-Plasma	1.97	0.522 0.635	0.001 0.000
Choline	3.21(s)	↑	1D-Plasma	3.22	0.548	0.001

Acetone	2.22(s)	↑	CPMG 1D-Plasma	2.23	0.338 0.370	0.043 0.026
D-3-hydroxybutyrate	1.20(d); 2.31(dd); 2.41(dd); 4.16(dt)	↑	CPMG 1D-Plasma	1.19	0.347 0.428	0.038 0.009
Acetoacetic acid	2.27(s); 3.43(s)	↑	CPMG 1D-Plasma	2.28	0.326 0.411	0.052 0.013
Glucose	3.244(dd); 3.402(m); 3.49(m); 3.535(dd); 3.728(m); 3.833(m); 3.898(dd); 4.647(d); 5.233(d)	↑	1D-Plasma	5.23	0.538	0.001
Scyllo-inositol	3.36(s)	↑	1D-Plasma	3.36	0.469	0.004
*Guanine	7.68(s)	↓	Urine	7.67	-0.535	0.002
Metabolite	**Chemical shifts (ppm) with multiplicities**	**Cholelithiasis vs. ethanol**	**Matrix**	**driver (δ)**	**Correl. (*r*)**	***p* value**
Acetone	2.22(s)	↓	Urine CPMG	2.24 2.23	-0.507 -0.472	0.077 0.065
Acetoacetic acid	2.27(s); 3.43(s)	↓	Urine CPMG 1D-Plasma	2.29 3.44 2.28	-0.456 -0.52 -0.453	0.118 0.039 0.078
3-hydroxyisovalerate	1.27 (s), 2.37(s)	↓	Urine	2.37	-0.851	0.000

Unknown	3.23(s)	↑	Urine	3.23	0.556	0.049
Metabolite	**Chemical shifts (ppm) with multiplicities**	**Glasgow score 1 vs. 2-3**	**Matrix**	**driver (δ)**	**Correl. (*r*)**	***p* value**
Lipid CH_3CH_2	0.93(m)	↓	1D-Plasma CPMG	0.925 0.925	-0.565 -0.560	0.023 0.024
Lipid $CH_2C=C$	2.00 (m)	↓	1D-Plasma	2.026	-0.554	0.026
Creatinine	3.05(s); 4.05(s)	↑	Urine	3.046	0.655	0.015
Malonic acid	3.12(s)	↑	Urine	3.119	0.625	0.023
Choline	3.21(s)	↑	CPMG	3.214	0.600	0.014
Valine	0.98(d); 1.03(d); 2.26(m); 3.60(d)	↑	CPMG	1.044	0.690	0.003
Isoleucine	0.94(t); 1.01(d); 1.26(m); 1.48(m); 1.98(m); 3.68(m)	↑	CPMG	1.012	0.550	0.027
Alanine	1.47(d); 3.77(q)	↑	CPMG	1.479	0.542	0.03
Acetyl signals from α1-acid glycoprotein (NAC1)	2.04(s)	↑	CPMG	2.037	0.557	0.025
Glucose	3.244(dd); 3.402(m); 3.49(m); 3.535(dd); 3.728(m); 3.833(m); 3.898(dd); 4.647(d); 5.233(d)	↓	1D-Plasma	3.244	-0.723	0.002
**Scyllo*-inositol	3.36(s)	↑	CPMG	3.355	0.569	0.021

*Guanine	7.68(s)	↑	Urine	7.679	0.657	0.015
Unknown	2.83(s)	↑	Urine	2.834	0.624	0.023

Key. Significance of the Pearson Correlation Coefficient (r) is given along with the p value for the specific driver (δ); * putative assignment.

Table 3. Statistically significant metabolites and their corresponding chemical shifts in the control comparison Cholelithiasis vs. Colonic inflammation.

Metabolite	Chemical shifts (ppm) with multiplicities	Colonic Inflammation	Matrix	driver (δ)	Correl. (*r*)	*p* value
Indole-3-acetate	3.65(s); 7.15(dd); 7.24(m); 7.50(d); 7.62(d)	↓	Urine	7.50	-0.711	0.004
Lactate	1.33(d); 4.11(q)	↓	CPMG	1.33	-0.623	0.017
		↓	1D-Plasma	1.32	-0.777	0.001
Lipid (mainly VLDL)	0.87(t)	↓	1D-Plasma	0.87	-0.635	0.015
Lipid (mainly LDL)	1.25(m)	↓	1D-Plasma	1.26	-0.702	0.005
Lipid (mainly VLDL)	1.56(m)	↓	1D-Plasma	1.56	-0.628	0.016
Lipid CH2CH=CH	2.00(m)	↓	1D-Plasma	2.00	-0.604	0.022
Leucine	0.95(t); 1.7(m); 3.72(m)	↓	1D-Plasma	0.95	-0.771	0.001
Valine	0.98(d); 1.03(d); 2.26(m); 3.60(d)	↓	1D-Plasma	1.03	-0.584	0.028
Malonic acid	3.12(s)	↑	Urine	3.12	0.652	0.012
Choline	3.21(s)	↓	1D-Plasma	3.22	-0.541	0.046
3-hydroxyvalerate	1.27 (s), 2.37(s)	↑	Urine	2.37	0.656	0.011

Key. Significance of the Pearson Correlation Coefficient (*r*) is given along with the *p* value for the specific driver (δ).Table 1. Summary of clinical data.

ASSOCIATED CONTENT

Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

ACKNOWLEDGMENTS

CEMBIO authors acknowledge the funding to Spanish Ministry of Economy and Competitiveness (MEC) grant CTQ2011-23562 and AV acknowledges her fellowship to EADS CASA and Santander bank for the mobility grant.

ABBREVIATIONS

AP, Acute pancreatitis; SIRS, Systemic Inflammatory Response Syndrome; USS, Abdominal ultrasound imaging; CT, Computed tomography; ^{1}H NMR, Proton Nuclear Magnetic Resonance; ERCP, Endoscopic Retrograde Cholangio Pancreatography; FID, Free Induction Decay; CPMG, Carr-Purcell-Meiboom-Gill; TSP, trimethylsilypropionate; PCA, Principal component analysis; PLSDA, Partial Least Square Discriminant Analysis; OPLSDA, Orthogonal-PLSDA; STOCSY, Statistical Total Correlation Spectroscopy; TOCSY, Total Correlation Spectroscopy; HMQC, Heteronuclear Multiple-Quantum Correlation; HSQC, Heteronuclear Single-Quantum Correlation; APACHE II, Acute Physiology And Chronic Health Evaluation II; LOS, length of stay; NSAID, Non-steroidal anti-inflammatory drug; TPR, True Positive Rates; FPR, False Positive Rates; ROC, receiver operating characteristic; AUC, area under the ROC curve; VLDL, Very-low-density lipoprotein; LDL, Low-density lipoprotein; CP, chronic pancreatitis; NAC1, acetyl signal from α1-acid glycoprotein, LC-MS, liquid chromatography coupled to mass spectrometry; PAG, phenylacetylglutamine.

REFERENCES

1. Baddeley, R. N.; Skipworth, J. R.; Pereira, S. P., Acute pancreatitis. *Medicine* **2010,** 39, (2), 108-115.

2. Fan, B.-G.; Andrén-Sandberg, Å., Acute pancreatitis. *North American Journal of Medical Sciences* **2010,** 2, (5), 211-214.

3. Treacy, J.; Williams, A.; Bais, R.; Willson, K.; Worthley, C.; Reece, J.; Bessell, J.; Thomas, D., Evaluation of amylase and lipase in the diagnosis of acute pancreatitis. *ANZ J Surg* **2001,** 71, (10), 577-82.

4. Xiao, B.; Zhang, X. M., Magnetic resonance imaging for acute pancreatitis. *World J Radiol* **2010,** 2, (8), 298-308.

5. Lempinen, M.; Puolakkainen, P.; Kemppainen, E., Clinical value of severity markers in acute pancreatitis. *Scand J Surg* **2005,** 94, (2), 118-23.

6. Jorge Cerrudo, J.; Mansilla Roselló, A.; Paz Yañez, A.; Segura Jiménez, I.; Ferrón Orihuela, J. A., Urgent vascular complications in acute pancreatitis. *Cir Esp* **2011**.

7. Raimondi, S.; Lowenfels, A. B.; Morselli-Labate, A. M.; Maisonneuve, P.; Pezzilli, R., Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. *Best Pract Res Clin Gastroenterol* **2010,** 24, (3), 349-58.

8. Vonlaufen, A.; Wilson, J. S.; Apte, M. V., Molecular mechanisms of pancreatitis: current opinion. *J Gastroenterol Hepatol* **2008,** 23, (9), 1339-48.

9. Chiang, D. T.; Anozie, A.; Fleming, W. R.; Kiroff, G. K., Comparative study on acute pancreatitis management. *ANZ J Surg* **2004,** 74, (4), 218-21.

10. Keller, J.; Layer, P., Idiopathic chronic pancreatitis. *Best Pract Res Clin Gastroenterol* **2008,** 22, (1), 105-13.

11. Bezmarević, M.; Kostić, Z.; Jovanović, M.; Micković, S.; Mirković, D.; Soldatović, I.; Trifunović, B.; Pejović, J.; Vujanić, S., Procalcitonin and BISAP score versus C-reactive protein and APACHE II score in early assessment of severity and outcome of acute pancreatitis. *Vojnosanit Pregl* **2012,** 69, (5), 425-31.

12. Sekimoto, M.; Takada, T.; Kawarada, Y.; Hirata, K.; Mayumi, T.; Yoshida, M.; Hirota, M.; Kimura, Y.; Takeda, K.; Isaji, S.; Koizumi, M.; Otsuki, M.; Matsuno, S.; JPN, JPN Guidelines for the management of acute pancreatitis: epidemiology, etiology, natural history, and outcome predictors in acute pancreatitis. *J Hepatobiliary Pancreat Surg* **2006,** 13, (1), 10-24.

13. Spanier, B. W.; Dijkgraaf, M. G.; Bruno, M. J., Epidemiology, aetiology and outcome of acute and chronic pancreatitis: An update. *Best Pract Res Clin Gastroenterol* **2008,** 22, (1), 45-63.

14. Nicholson, J. K.; Wilson, I. D., Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* **2003,** 2, (8), 668-76.

15. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R., Metabonomics: Metabolic Processes Studied by NMR Spectroscopy of Biofluids. *Concepts in Magnetic Resonance* **2000,** 12, (5), 289-320.

16. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *J Proteome Res* **2007,** 6, (2), 443-58.

17. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* **2002,** 1, (2), 153-61.

18. Lindon, J. C.; Nicholson, J. K., Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *Trends in Analytical Chemistry* **2008,** 27, (3), 194-204.

19. Lusczek, E. R.; Paulo, J. A.; Saltzman, J. R.; Kadiyala, V.; Banks, P. A.; Beilman, G.; Conwell, D. L., Urinary ^{1}H-NMR metabolomics can distinguish pancreatitis patients from healthy controls. *JOP* **2013,** 14, (2), 161-70.

20. UK guidelines for the management of acute pancreatitis. *Gut* **2005,** 54 Suppl 3, iii1-9.

21. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* **2007,** 2, (11), 2692-703.

22. Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K., Recursive segment-wise peak alignment of biological ^{1}H NMR spectra for improved metabolic biomarker recovery. *Anal Chem* **2009,** 81, (1), 56-66.

23. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Anal Chem* **2006,** 78, (13), 4281-90.

24. Eriksson, I.; Johansson, E.; Kettaneh-Wold, N.; Wold, S., *Multi-and Megavariate Data Analysis and Applications.* Umetrics: Umea: 2001; p 533.

25. Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J.; Holmes, E.; Trygg, J., OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics* **2006,** 20, 341–351.

26. Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J., Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic ^{1}H NMR data sets. *Anal Chem* **2005,** 77, (5), 1282-9.

27. Cartwright, S. L.; Knudson, M. P., Evaluation of acute abdominal pain in adults. *Am Fam Physician* **2008,** 77, (7), 971-8.

28. Bales, J. R.; Sadler, P. J.; Nicholson, J. K.; Timbrell, J. A., Urinary excretion of acetaminophen and its metabolites as studied by proton NMR spectroscopy. *Clin Chem* **1984,** 30, (10), 1631-6.

29. Spraul, M.; Hofmann, M.; Dvortsak, P.; Nicholson, J. K.; Wilson, I. D., High-performance liquid chromatography coupled to high-field proton nuclear magnetic resonance spectroscopy: application to the urinary metabolites of ibuprofen. *Anal Chem* **1993,** 65, (4), 327-30.

30. Garcia-Perez, I.; Couto Alves, A.; Angulo, S.; Li, J.; Utzinger, J.; Ebbels, T.; Legido-Quigley, C.; Nicholson, J.; Holmes, E.; Barbas, C., Bidirectional correlation of NMR and

capillary electrophoresis fingerprints: a new approach to investigating Schistosoma mansoni infection in a mouse model. *Anal Chem* **2010,** 82, (1), 203-10.

31. Gohel, M. C.; Jogani, P. D., A review of co-processed directly compressible excipients. *J Pharm Pharm Sci* **2005,** 8, (1), 76-93.

32. Zheng, X.; Xie, G.; Zhao, A.; Zhao, L.; Yao, C.; Chiu, N. H.; Zhou, Z.; Bao, Y.; Jia, W.; Nicholson, J. K., The footprints of gut microbial-mammalian co-metabolism. *J Proteome Res* **2011,** 10, (12), 5512-22.

33. Lees, H. J.; Swann, J. R.; Wilson, I. D.; Nicholson, J. K.; Holmes, E., Hippurate: The Natural History of a Mammalian-Microbial Cometabolite. *J Proteome Res* **2013**.

34. Coleman, M. D.; Norton, R. S., Observation of drug metabolites and endogenous compounds in human urine by 1H nuclear magnetic resonance spectroscopy. *Xenobiotica* **1986,** 16, (1), 69-77.

35. Kinross, J. M.; Darzi, A. W.; Nicholson, J. K., Gut microbiome-host interactions in health and disease. *Genome Med* **2011,** 3, (3), 14.

36. Wu, D.; Xu, Y.; Zeng, Y.; Wang, X., Endocrine pancreatic function changes after acute pancreatitis. *Pancreas* **2011,** 40, (7), 1006-11.

37. Stefanutti, C.; Labbadia, G.; Morozzi, C., Severe hypertriglyceridemia-related acute pancreatitis. *Ther Apher Dial* **2013,** 17, (2), 130-7.

38. Serpytis, M.; Karosas, V.; Tamosauskas, R.; Dementaviciene, J.; Strupas, K.; Sileikis, A.; Sipylaite, J., Hypertriglyceridemia-induced acute pancreatitis in pregnancy. *JOP* **2012,** 13, (6), 677-80.

39. Khovidhunkit, W.; Kim, M. S.; Memon, R. A.; Shigenaga, J. K.; Moser, A. H.; Feingold, K. R.; Grunfeld, C., Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. *J Lipid Res* **2004,** 45, (7), 1169-96.

40. Bartolomé, N.; Aspichueta, P.; Martínez, M. J.; Vázquez-Chantada, M.; Martínez-Chantar, M. L.; Ochoa, B.; Chico, Y., Biphasic adaptative responses in VLDL metabolism and lipoprotein homeostasis during Gram-negative endotoxemia. *Innate Immun* **2012,** 18, (1), 89-99.

41. Solanki, N. S.; Barreto, S. G.; Saccone, G. T., Acute pancreatitis due to diabetes: the role of hyperglycaemia and insulin resistance. *Pancreatology* **2012,** 12, (3), 234-9.

42. Fang, F.; He, X.; Deng, H.; Chen, Q.; Lu, J.; Spraul, M.; Yu, Y., Discrimination of metabolic profiles of pancreatic cancer from chronic pancreatitis by high-resolution magic angle spinning ^{1}H nuclear magnetic resonance and principal components analysis. *Cancer Sci* **2007,** 98, (11), 1678-82.

43. Adrych, K.; Smoczynski, M.; Stojek, M.; Sledzinski, T.; Slominska, E.; Goyke, E.; Smolenski, R. T.; Swierczynski, J., Decreased serum essential and aromatic amino acids in patients with chronic pancreatitis. *World J Gastroenterol* **2010,** 16, (35), 4422-7.

44. Ma, C.; Tian, B.; Wang, J.; Yang, G. J.; Pan, C. S.; Lu, J. P., Metabolic characteristics of acute necrotizing pancreatitis and chronic pancreatitis. *Mol Med Rep* **2012,** 6, (1), 57-62.

45. Zhang, L.; Jin, H.; Guo, X.; Yang, Z.; Zhao, L.; Tang, S.; Mo, P.; Wu, K.; Nie, Y.; Pan, Y.; Fan, D., Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by ^{1}H nuclear magnetic resonance-based metabonomic profiles. *Clin Biochem* **2012,** 45, (13-14), 1064-9.

46. Wang, J.; Ma, C.; Liao, Z.; Tian, B.; Lu, J. P., Study on chronic pancreatitis and pancreatic cancer using MRS and pancreatic juice samples. *World J Gastroenterol* **2011,** 17, (16), 2126-30.

47. Giakoustidis, A.; Mudan, S. S.; Giakoustidis, D., Dissecting the stress activating signaling pathways in acute pancreatitis. *Hepatogastroenterology* **2010,** 57, (99-100), 653-6.

48. Nicholas, P. C.; Kim, D.; Crews, F. T.; Macdonald, J. M., ^{1}H NMR-based metabolomic analysis of liver, serum, and brain following ethanol administration in rats. *Chem Res Toxicol* **2008,** 21, (2), 408-20.

49. Zuo, Y. Y.; Kang, Y.; Yin, W. H.; Wang, B.; Chen, Y., The association of mean glucose level and glucose variability with intensive care unit mortality in patients with severe acute pancreatitis. *J Crit Care* **2012,** 27, (2), 146-52.

50. Billingham, M. E.; Gordon, A. H., The role of the acute phase reaction in inflammation. *Agents Actions* **1976,** 6, (1-3), 195-200.

51. Smelt, A. H., Triglycerides and gallstone formation. *Clin Chim Acta* **2010,** 411, (21-22), 1625-31.

52. Nishijima, T.; Nishina, M.; Fujiwara, K., Measurement of lactate levels in serum and bile using proton nuclear magnetic resonance in patients with hepatobiliary diseases: its utility in detection of malignancies. *Jpn J Clin Oncol* **1997,** 27, (1), 13-7.

53. Kobori, K.; Sakakibara, H.; Maruyama, K.; Kobayashi, T.; Yamaki, T., [A rapid method for determining urinary indoleacetic acid concentration and its clinical significance as the tumor-marker in the diagnosis of malignant diseases]. *J UOEH* **1983,** 5, (2), 213-20.

54. Kelly, D. A., Intestinal failure-associated liver disease: what do we know today? *Gastroenterology* **2006,** 130, (2 Suppl 1), S70-7.

55. Balasubramanian, K.; Kumar, S.; Singh, R. R.; Sharma, U.; Ahuja, V.; Makharia, G. K.; Jagannathan, N. R., Metabolism of the colonic mucosa in patients with inflammatory bowel diseases: an in vitro proton magnetic resonance spectroscopy study. *Magn Reson Imaging* **2009,** 27, (1), 79-86.

56. Suvorova, I. A.; Ravcheev, D. A.; Gelfand, M. S., Regulation and evolution of malonate and propionate catabolism in proteobacteria. *J Bacteriol* **2012,** 194, (12), 3234-40.

SUPPORTING INFORMATION.

Figure 1- Supplemental. STOCSY plot derived from the correlation matrix calculated between the data point at the peak of 3-hydroxyisovalerate at δ 2.37 and all other data points (matrix: urinary ^{1}H-NMR spectra, comparison: diagnosis of AP). Medium correlation (r=0.6) is shown in yellow/green peak with resonances from 2.17(s).

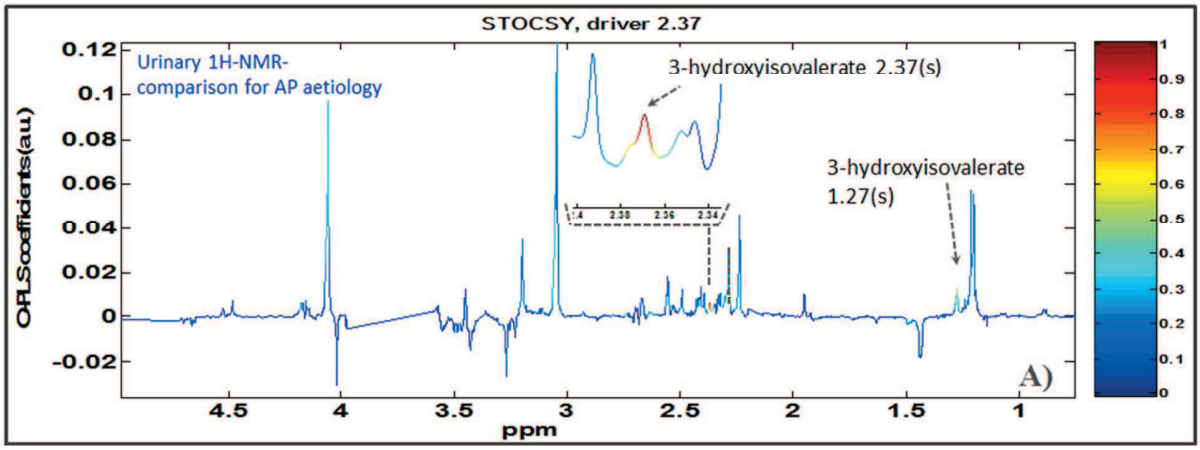


Figure 2-Supplemental. [1]H-NMR spectra from a patient with acute on chronic alcoholic pancreatitis. Spectra shows the pathological condition and broadened spectrum can be observed, these probably caused by proteins in urine therefore it is not suitable for metabonomic approachs. Key: 1. Creatinine [δ (ppm): 3.05(s); 4.06(s)]; 2. Acetaminophen glucuronide (AG) [δ (ppm): 2.17(s), 3.62(m), 5.10(d), 7.13(d), 7.36(d)]; 3. Acetaminophen (A) [δ (ppm): 2.16(s), 6.91(d), 7.25(d)]; 4. Acetaminophen sulphate (AS) [δ (ppm): 2.18(s), 7.31(d), 7.46(d)]; 5. Hippurate [δ (ppm): 3.97(d); 7.56(t); 7.64(t); 7.84(d)].

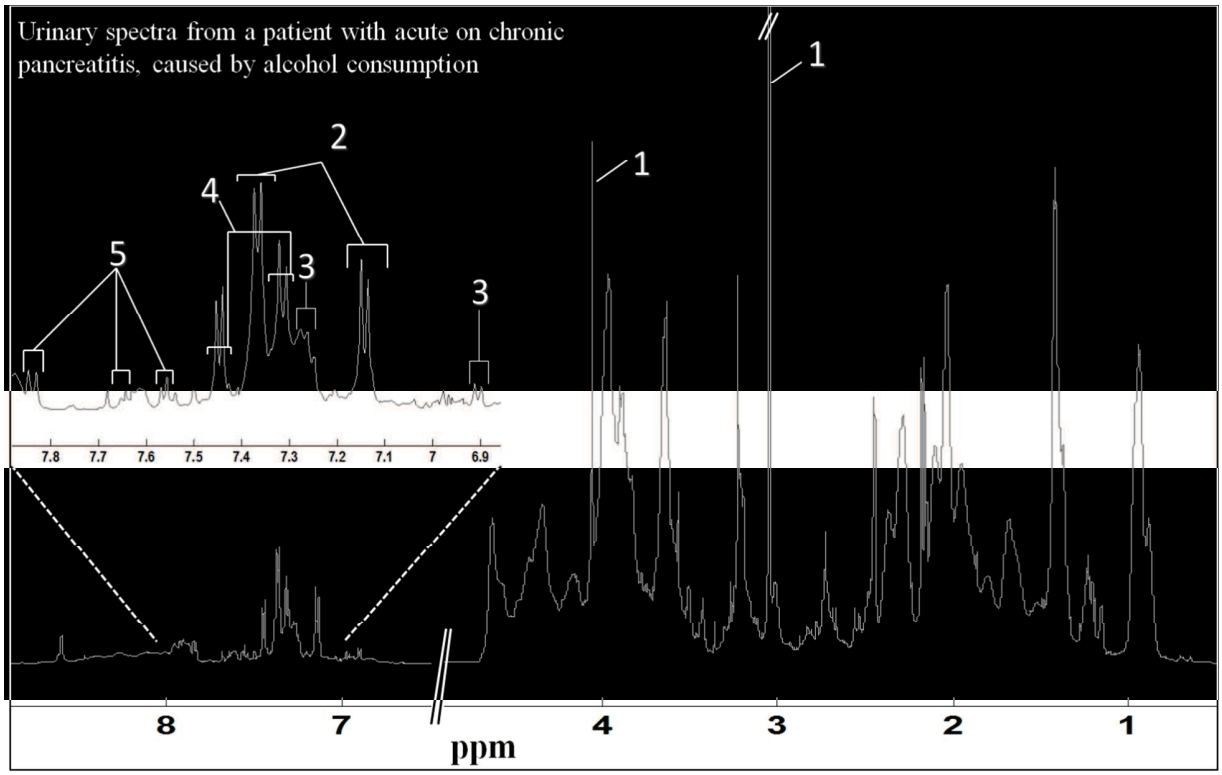


Figure 3-Supplemental. Pearson correlation structure obtained after selecting driver peak from mannitol [δ) 3.805, (5-CH)] in the 1D NMR profile and applying a p=0.05 significance criterion. Signals positively and highly correlated to mannitol signal can be seen highlighted in red chemical shifts, such as mannitol [signals on NMR; doublet of doublets δ) 3.69, (r) 0.985, p_value) 0.002; multiplet δ) 3.77, (r) 0.991, p_value) 0.001; doublet δ)3.81, (r) 0.992, p_value) 0.0009; doublets of doublets δ) 3.88, (r) 0.991, p_value) 0.001], Acetaminophen Glucuronide (AG) [singlet δ) 2.16, (r) 0.999, p_value) $9.84E^{-05}$; multiplet δ) 3.62, (r) 0.960, p_value) 0.009; doublet δ) 7.15, (r) 0.998, p_value) 0.0001; doublet δ) 7.36, (r) 0.992, p_value) 0.0009], Acetaminophen (A) [singlet δ) 2.15, (r) 0.956 p_value) 0.0009; doublet δ) 6.95, (r) 0.976, p_value) 0.004; doublet δ) 7.25, (r) 0.957, p_value) 0.011], Acetaminophen Sulfate (AS) [singlet δ) 2.18, (r) 0.955 p_value) 0.011; doublet δ) 7.32, (r) 0.941, p_value) 0.017; doublet δ) 7.44, (r) 0.922, p_value) 0.026], Acetylcysteine conjugate of acetaminophen (NAC) [singlet δ 1.86, (r) 0.902 p_value) 0.036; singlet δ 2.15, (r) 0.926, p_value) 0.024; doublet δ 6.98, (r) 0.897, p_value) 0.039].

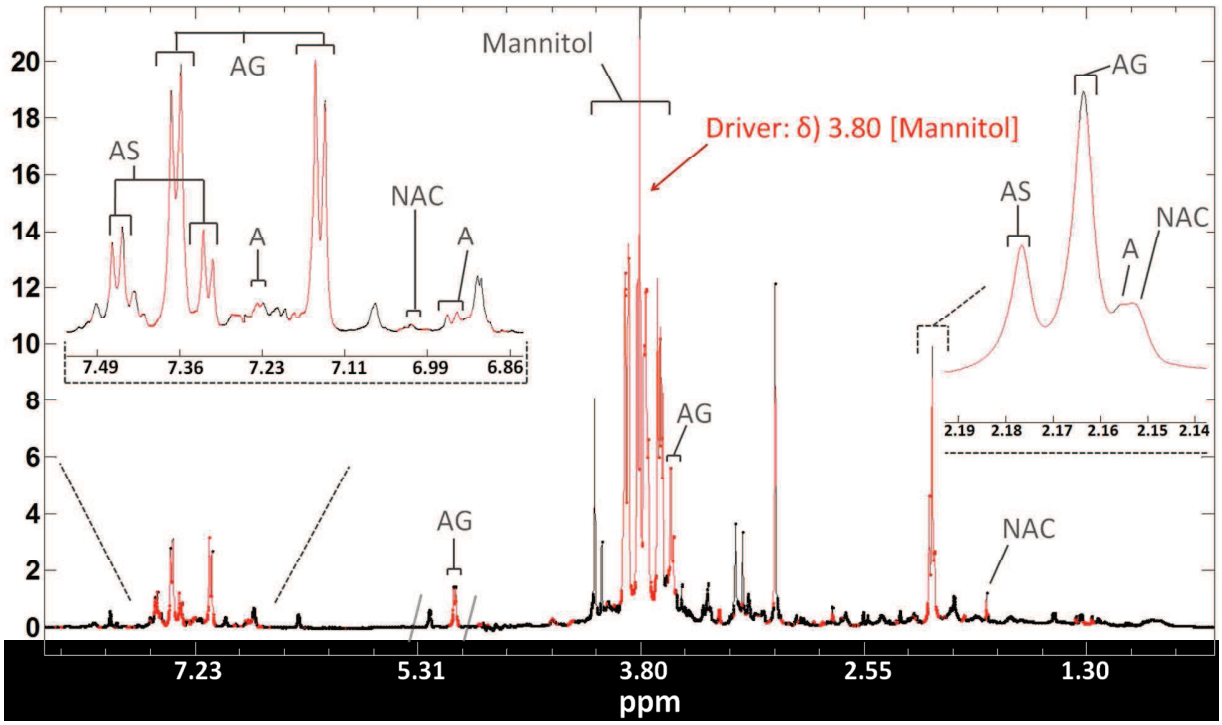


Figure 4- Supplemental. 600 MHz CPMG spectra (δ -0.0 to 4.7) of serum plasma: (A) acute pancreatitis patient, (B) Control patient with diverticulitis. Abbreviations: NAC1 and NAC2 refer to composite acetyl signals from α1-acid glycoprotein.

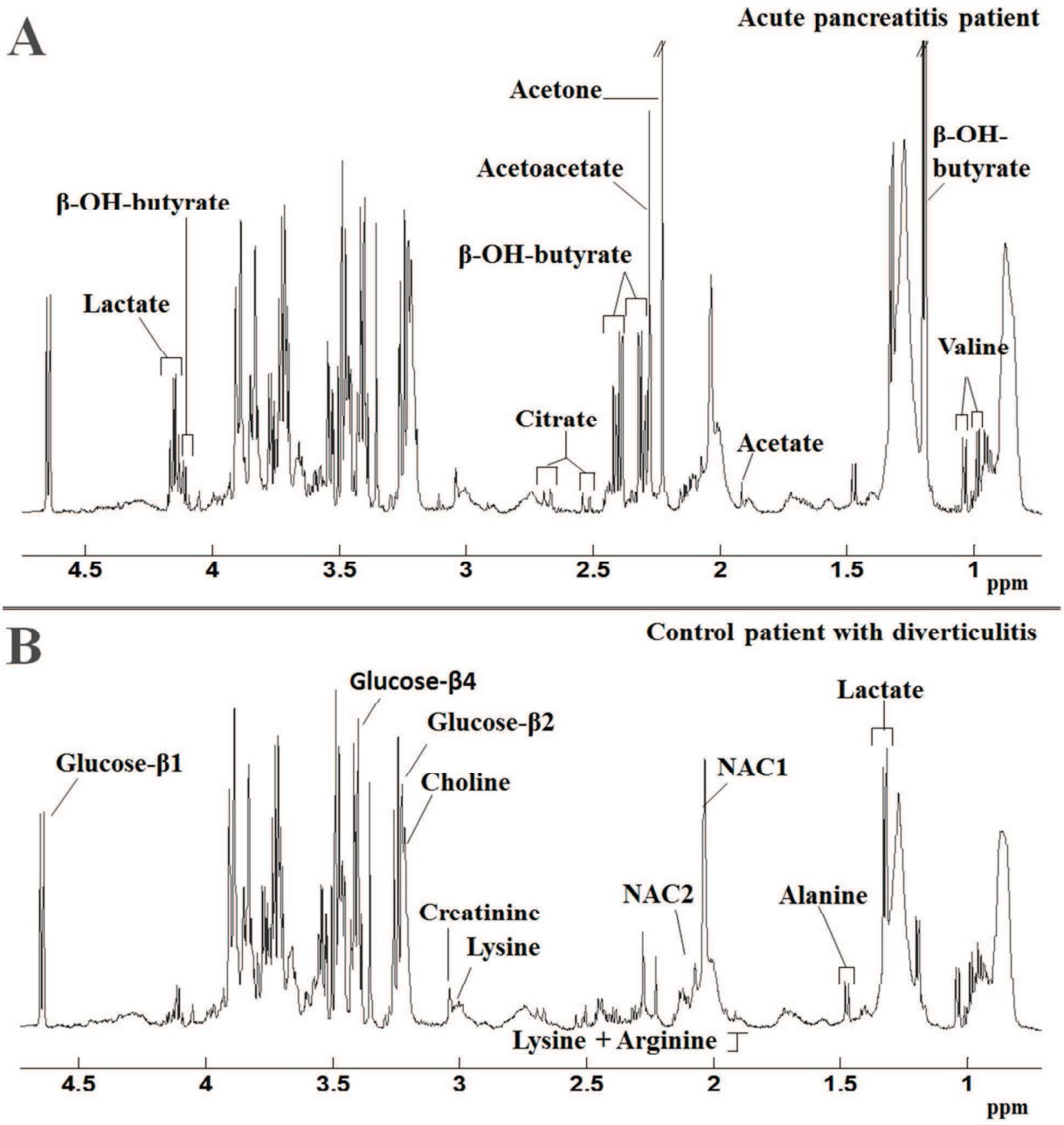


Figure 5- Supplemental. STOCSY plot derived from the correlation matrix calculated between the data point at the peak of Acetaminophen at δ 2.16 and all other data points (matrix: urinary ^{1}H-NMR spectra). Strong correlation is shown in red/orange color. High and positive correlation is observed with all acetaminophen related metabolites(A, AG, AS and NAC).

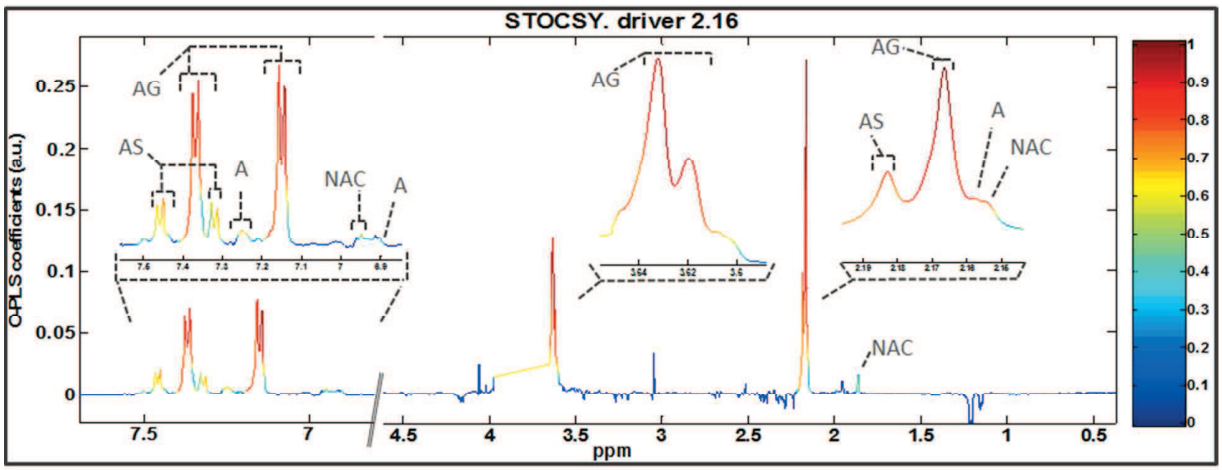


Table 1- Supplemental. Statistically significant metabolites and their corresponding chemical shifts in the antibiotic consumers group vs controls.

Metabolite	**Coupling pattern (Hz)**	**Antibiotic consumers**	**driver (δ)**	**Correl. (*r*)**	***p* value**
Creatinine	3.05(s); 4.06(s)	↓	3.047	-0.506	0.004
3-hydroxyisovaleric acid	1.273(s); 2.35(s)	↓	1.273	-0.536	0.002
Methylamine	2.56(s)	↓	2.550	-0.541	0.002
Citrate	2.68(d); 2.54(d)	↓	2.693	-0.471	0.008
Dimethylglycine	2.95(s); 3.71(s)	↓	2.946	-0.500	0.004
Unknown [2-ketobutyric acid]	1.07(t); 2.76(q)	↓	1.040	-0.543	0.002
Unknown [Metronidazole]	4.61(t)	↑	4.613	0.603	0.000
Unknown	1.95(s)	↑	1.954	0.371	0.040
Unknown	4.02(s); 6.66(m); 6.91(d); 7.71(s)	↑	6.658	0.482	0.006

Key. Significance of the Pearson Correlation Coefficient (*r*) is given along with the *p* value for the specific driver (δ). The compounds which are mentioned above within brackets are proposed or tentative identifications.

Table 2- Supplemental. Biochemical clinical parameters a comparison between AP and Control group.

	Control Patients			AP Patients			*p value*
	Number	**Average**	**SD**	**Number**	**Average**	**SD**	
Cases AP/Control	23			15			
Body temperature (°C)	22	37.2	0.6	13	36.7	0.4	0.01*
SBP (mmHg)	22	122.8	15.2	14	149.9	16.8	4.9E-05**
DBP (mmHg)	22	71.8	9.7	14	88.4	10.5	2.8E-05**
pH	6	7.4	0.0	13	7.4	0.1	0.72
PaO2 (mmHg)	6	9.4	3.1	13	11.7	2.4	0.09
PaCO2 (mmHg)	6	4.8	0.1	13	4.5	0.7	0.10
Amylase (U/L)	21	62.0	73.7	15	874.6	841.8	1.3E-06**
Amylase 2 (U/L)	8	67.3	77.8	13	325.4	210.8	1.3E-03**
WCC (cells/uL)	23	12.6	5.7	15	12.7	3.9	0.72
Hb (g/dL)	23	14.1	1.5	15	14.9	0.9	0.07
Hct (%)	23	0.4	0.0	15	0.5	0.1	0.10
Na (mEq/L)	23	136.3	2.7	15	136.3	5.3	0.56
K (mEq/L)	23	3.9	0.4	15	3.8	0.6	0.41
Urea (mg/dL)	23	3.9	1.4	15	6.1	4.8	0.11
Creatinine (mg/dL)	23	82.6	12.9	15	140.6	174.9	0.22
Calcium (mg/dL)	22	2.3	0.1	15	2.3	0.2	0.51
Glucose (mg/dL)	11	5.9	1.5	12	9.0	4.0	0.03*
Bilirubin (mg/dL)	22	24.0	24.7	15	31.60	26.27	0.13
Albumin (g/dL)	22	42.8	4.6	15	41.7	5.2	0.74
ALP (IU/L)	22	133.0	293.1	15	99.8	38.6	0.05
ALT (IU/L)	20	84.5	189.0	15	126.0	122.0	0.03*
CRP (mg/L)	23	63.9	81.5	15	39.5	60.7	0.16

Data presented is average ± SD. Kolmogorov-Smirnov-Lilliefors test was applied for each parameter in order to assure normality of each measure, parameters that follow normality, were analyzed by unpaired two-tailed Student's t-test. Parameters without normality were analyzed using Mann–Whitney U test. Key: * $p<0.05$; ** $p<0.01$. Key: CRP [C-reactive protein]; ALT [alanine Aminotransferase]; ALP [Alkaline phosphatase]; K [Potassium]; Na [Sodium]; Hct [Haematocrit]; Hb [Haemoglobin]; WCC [White cell count]; PaCO2 [Partial Pressure of carbon dioxide]; PaO2 [Partial pressure of oxygen]; DBP [Diastolic Blood Pressure]; SBP [Systolic Blood Pressure].

CHAPTER 4

3. Capillary Electrophoresis–Ultraviolet detection (CE–UV) as a complementary technique of Nuclear Magnetic Resonance spectroscopy (NMR)

3.1. Fundamentals of the technique

Capillary electrophoresis (CE) is one of the most powerful separation techniques, specifically characterized by high resolution and separation efficiency. CE is based on the difference in mobility that molecules exhibit in an applied electric field, and how these are separated on their charge–to–mass ratio[1–3]. Historically, the first electrophoretic separation was published in the 1930s by Arne Tiselius, who, for this and other accomplishments, was awarded the Nobel Prize in 1948[2]. CE was developed in the 1980s by James Jorgenson and Krynn Lucas, which was automated and became a powerful tool for separation and purification. It is now a fast–growing method in academia and industry[2]. Despite CE being the last technique to be included in metabolomics, this had high improvement since then because the main components of biofluids display high polarity and are water soluble[1].

A typical CE instrumental diagram is shown in Figure 4.1. Here, the capillary is filled with buffer solution before the sample is injected. For separation, the capillary is dipped back to the buffer vial and then a high voltage is applied. The components of the sample will start to migrate towards the electrode at the destination vial. Migration of the compounds will be at different velocities according to their electrophoretic mobility (normally based on its charge and molecular size) and the velocity of the electro–osmotic flow (EOF). The capillary has a detection window towards its outlet that is connected to a computer. The compounds of the sample are plotted as peaks in an electropherogram[2]. In the case of CE coupled to MS, the CE system does not possess an outlet vial, and the OEF is directed toward the interface with the MS detector[3–4].

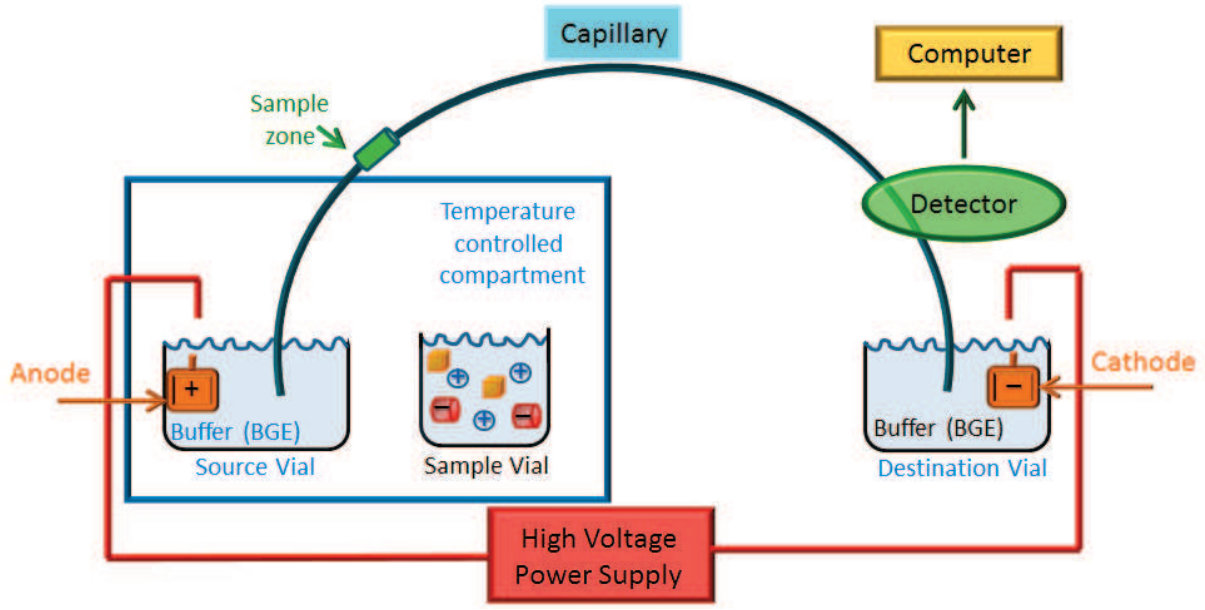


Figure 4.1. Typical capillary electrophoresis–instrumental diagram (Adapted from University of California, Davis, USA, webpage: http://chemwiki.ucdavis.edu/).

Different detection systems have been employed for CE, including ultraviolet (UV) absorption, fluorescence (e.g. LIF detector), electrochemical detection (ED) and MS[2, 5–6]. With each of these detectors CE has been used in the field of metabolomics, and examples are found more in detail elsewhere[5]. In comparison with other detection techniques, MS is universal. In addition, in metabolomics, MS is now the most widely used detector used coupled with CE, based on the capacity of MS for compound identification. However, other detectors such as CE–NMR have been reported for the analysis of biological samples. CE has also been used for multi–dimensional separation systems e.g. two–dimensional LC–CE and lastly, in the promising development of CE on a microchip[5]. Details and characteristics of CE–LIF are going to be discussed in chapter 6 of this thesis.

There are different separation modes used in CE (Table 4.1); the main ones for metabolomics applications are CZE and MEKC. Of the two, CZE is probably more used, and this has a wide variety of applications for small molecules and proteins[1, 7]. MEKC has great utility in separating mixtures that contain both ionic and, more importantly, neutral species, and this has become valuable in the separation

of very hydrophobic pharmaceuticals from their very polar metabolites[2]. Optimizing parameters in separations by CE are: pH; buffer type; temperature; and the addition of some organic solvents, interacting agents or surfactants. Table 4.1 summarizes the principal characteristics of these modes; the separation features and the principles are also presented.

Table 4.1 Characteristics of various CE modes (adapted from Ahmed, F., 2009[7]).

Mode	Separation feature	Principle	Characteristics
Capillary Zone Electrophoresis (CZE)	▪ Charge–to–size ratio	▪ pH of the running buffer determines effective analyte charge	▪ Stable & constant EOF ▪ EOF dominates the separation process then analyte electropheric mobility ▪ Normal migration order: cations> neutrals> anions ▪ Neutrals do not separate
Micellar Electrokinetic Chromatography (MEKC)	▪ Electrophoretic and electroosmotic mobility of ion pairs with opposite charges in EOF–driven system	▪ Surfactants (e.g. anionic SDS) will form charged micelles which are added allowing separation of neutral analytes.	▪ Differences in migration are mainly due to variation in the distribution of uncharged analytes across the differing degrees of micelles ▪ EOF dominates the separation process ▪ Possible to add methanol and cyclodextrins to improve solubility and separation efficiency
Capillary Isoelectric Focusing (CIEF)	▪ Isoelectric point	▪ Running buffer containing ampholytes generating an electric–stable pH gradient	▪ It is often used for the separation of complex mixtures of peptides and proteins

Capillary Gel Electrophoresis (CGE)	▪ Size	▪ Gels and sieving polymers form a net that permits size–exclusion separation	▪ Faster, better separation efficiency, fully automated and improved sensitivity than typical gel technology ▪ Applied in the separation of DNA and RNA
Capillary Electrochroma–tography (CEC)	▪ Hybrid of chromatography & EOF	▪ Chromatographic separation combined with an electrostatically driven separation	▪ Packed columns are mostly employed for separation, although open tubular formats gained wider use because of their simplicity and the possibility of commercialization
Capillary Isotachophoresis (CITP)	▪ Mobility	▪ Discontinuous buffer system through the run creates different separation zones	▪ Upon the application of voltage, the sample components migrate into separate but adjoining zones according to their electrophoretic mobilities

3.2. Characteristics of CE–UV for metabolomics studies

CE is an analytical separation technique capable of high–resolution separation of a diverse range of chemical compounds, and is particularly suitable for the separation of polar and charged compounds. Some characteristics of CE are: high resolution and peak capacity; high efficiency; selective through multiple separation modes; relatively short time of analysis; requires small amounts of samples, background solutions, and reagents if they are needed. CE has the ability to solve complex sample matrix effects through simple modifications of the electrolyte system[8–9]. One of the important characteristics of CE is the low organic solvent consumption (sometimes even none) which means it can be considered a green technology[1, 3, 5]. In addition, CE–UV has a low cost compared to all the high–throughput technologies used in metabolomics.

One of the greatest benefits of CE is the ability to separate small compounds in complex matrices with almost no sample preparation: e.g. urine can be injected directly. Also, after each analysis the capillary can easily be washed, reducing the capillary blockage by particles. In case of samples with high protein content (e.g. serum or cerebrospinal fluid), deproteinization is often applied to prevent precipitation or adsorption of proteins in the capillary[1, 5–6]. These characteristics make it a powerful complement or alternative technique to be used in metabolomics.

The major drawback of CE is its poor sensitivity based on the limited amount of sample volume that can be introduced into the capillary (nanoliters), and the low absorption due to the short optical path–length if a UV detector is used. The UV detector at low wavelengths is a non–selective detection method i.e. it has the potential to detect nearly every chemical compound which, for non–target analysis, is an advantage. However compounds without chromophores (e.g. carbohydrates) are difficult to detect by UV[1, 3, 5]. Otherwise, higher sensitivity and lower limits of detection are obtained when CE is coupled with MS, which also has the advantage of providing structural information. Unfortunately, there are some critical aspects associated with coupling CE to MS. The number of ionization techniques that can be used is limited i.e. ESI is the most commonly used for interface. Also, most of CE buffers may be incompatible with the ion source, and consequently only volatile and pure buffers are used (ammonium, acetate or formic acid)[1, 5]. Furthermore, CE–MS coupling is difficult because the very low flow rates of CE cause the electrospray to be unstable. This was overcome by the use of a sheath–flow interface, which had a negative effect on sensitivity. Low sensitivity was also improved by using sample pre–treatment techniques to concentrate the analytes (such as solid–phase extraction, SPE)[6]. Nevertheless, the simplest and most used on–line pre–concentration procedure is sample stacking. In this method, a sample plug is introduced with a buffer with a higher pH than the separation buffer[7].

Another problem of UV detection is the identification; this has to be done by spiking the sample with the standard. This process must be done via trial by error assay with pure standards of compounds expected in the profile. The assignment is performed by comparison of migration times, absorption spectrum, and by the increase of the peak signal due to the standard[8]. Even for CE–MS (where TOF is the typical mass analyser) there are no general mass–spectra databases of endogenous metabolites[5]. Although Sugimoto et al. reported a method to identify unknown peaks based on the predicted migration time (tm)

and accurate *m/z* values to assign tentative identities, the list of metabolites is not publicly available[1]. This turns data analysis and identification into a very time consuming process.

In addition, robustness is a weak point of the technique. This means that, in practice migration time shifts complicate serial analysis[1]. The reproducibility and shift of retention or migration time among the different separation techniques would be GC>LC>CE. Table 4.2 summarizes the capabilities and limitations of the CE–UV technique.

Table 4.2 Relative strengths and weakness of CE–UV for metabolomics studies[1].

Strengths	Drawbacks
▪ Suitable for polar and charged compounds	▪ Poor sensitivity and low limit of detection (µM)
▪ High resolution, peak capacity & high efficiency	▪ Migration time shifting
▪ Selective through multiple separation modes	▪ Difficult to assign identification (standard spike)
▪ Small amount of sample (few nanolitres) & minimum preparation for polar matrix (e.g. urine or saliva)	▪ Not universal detector (only compounds with UV chromophore)
▪ Low consumption of organic solvent & reagents if needed	▪ Reproducibility problem
▪ Low cost & simple instrumentation	
▪ UV detector at low wavelengths is a non–selective	

UV detection is often underestimated based on its poor sensitivity and selectivity, and because it does not provide structural information. Nevertheless, in third world countries, CE–UV could represent an emergent diagnostic tool for the most common diseases. CE–UV has proved to be an inexpensive tool able to provide profiles that permit classification of samples in several cases of study, such as: the determination of organic acidurias in neonates[10]; in the research of biomarkers for parasite diseases i.e. *schistosomasis*[11]; or in studying the effect of antioxidant diets on children with diabetes 1[12]. CE–UV could

have a cheap clinical application once the method could be exhaustively validated in humans. To accomplish this goal, it is necessary to enlist help from wealthy countries e.g. in the research of diagnostic compounds that later could be validated by CE–UV.

3.3. Insight into the application of CE–UV as a complementary technique to NMR

The development of CE–UV methodologies that could be suitable and answer the requirements of reproducibility, reliability and robustness in metabolomics investigations, were initiated with a method in CZE mode by Garcia, A., in 1998[10]. Then, a pioneering work by Guillo, C., in 2004[13] showed how sulphated β–CD–modified MEKC can be used to provide a useful tool for non–target analysis in urine. Both accomplished and presented recently as a new strategy that combines two methods operating in opposite polarities, and in this way exploited the separation mechanisms from both CE separation modes—namely CD–MEKC (with normal polarity) and CZE (with reversed polarity)[6, 8]. Since then, these methodologies have been applied in our centre for several studies in metabolomics. Barbas. C., in 2008 showed that CE separations performed with reverse polarity could provide the separation of over 60 signals in urine samples in around 25min. It's important to note that both methods confirm the possibility of measuring the same samples in CE operating, with each methodology (with different polarity and buffer system) affording the exciting opportunity of having a truly holistic metabolomics approach, since all compounds in a given sample can then be resolved and detected in one or other mode, with the only limitation being UV sensitivity[1]. For a normal polarity CD–MEKC method, cations and neutral compounds migrate first, followed by only large anions or weak acids at the end of the profile. Moreover, reverse polarity CZE method detects mainly ions, such as short chain organic acids, which migrate first from smaller to higher *m/z*.

These methodologies have strong advantages to be used: both are easily to apply; the buffer preparation is easy and cheap; the time of analysis is relatively short; and the migration time shift is low because the current generated by the capillary is around –70μA for CZE and 150μA for CD–MECK, which means no high current and so no increase of temperature in the system, despite the temperature control system. As expected, higher migration time shifting is observed for CD–MECK than CZE

method. In addition, once the samples are measured in both CE–modes, the combination of the two polarities profiles can be achieved (the specific process is detailed in this paper). In this way hundreds of metabolites from cations, neutrals and anions are displayed in one single set. For this particular paper, a metabolic phenotyping of urine using CD–MEKC and CZE mode, 33 compounds were identified and are shown in Figure 4.2A. As it can be seen in the paper, CE–UV like NMR has the potential of displaying significant differences in the peaks by visual inspections of the profile among the groups involved in a study (Figure 4.2B).

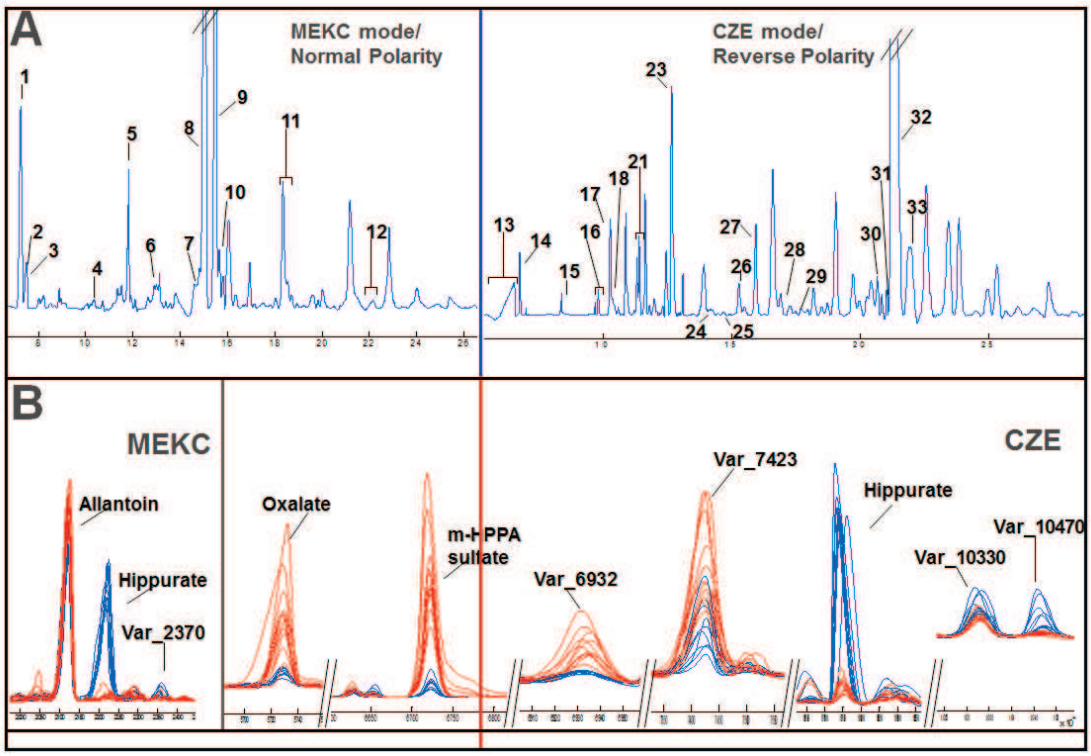


Figure 4.2. (A) CE–UV metabolic fingerprinting of urine corresponding to male wild type mouse, and (B) different levels observed in some metabolites according to the knockout condition. ID: 1. Urea; 2. Creatinine; 3. Creatine; 4. Histidine; 5. Phenylalanine; 6. Methyluridine; 7. PAG; 8. Allantoin; 9. Hippurate; 10. Uridine; 11. Uric/Inosine; 12. Benzoate; 13. Chlorides; 14. Nitrate; 15. Oxalate; 16. Fumarate; 17. Oxoglutarate; 18. Succinate; 19. Isocitrate; 20. Glutarate; 21. Citrate; 22.m–HPPA sulfate (theorical); 23. Glycolate (coelutes); 24. Acetoacetate; 25. Lactate; 26. Ketosiovalerate; 27. Glycerate/Benzoate; 28. α–Ketoisocaproate; 29. Phenylpyruvate; 30. Phenyllactate; 31. Glutamate; 32. Hippurate/ Vanillate 33. Ascorbate. Wild type control highlighted in blue. KO mice highlighted in red. Figure taken from[14].

As a short background to this project; the deletion of slc26a6 (CFEX) gene in mice would have awakened the interest of the researchers because they clearly visualized the formation of oxalate stones in the kidney. This made the urinary metabolic characterization an interesting approach for the knockout model. So a dual strategy was planned, combining NMR spectroscopy with CE–UV for the same urine sample to comprehensively describe the urinary metabolic phenotype of the CFEX–null mouse model.

Both technologies have the characteristic of producing 2D data matrices and so the data treatment is similar. In addition, both have been applied to the non–targeted analysis of various diseases, and more importantly together offer complementary information. While NMR generates comprehensive and robust profiles of a broad range of low molecular weight molecules, CE–UV can provide quantitative data on ionic species including oxalate, (which is not detected by NMR) chloride and nitrates. The combination of NMR and CE–UV profiles is described in the paper.

Very important to highlight is the fact that for such a small sample volume (around microliters for mice urine), this type of technique CE–UV can work fine, i.e. in total only 15µL of urine was employed just for CE-UV. Initially, this volume was mixed with hydrochloric acid (1:1) to be run in CZE method. Afterwards, the same samples were mixed with buffer and water for CD–MEKC method. This allowed for sample measurement in NMR (through the use of microprobes) and CE–UV in both modes.

The combination and the data treatment of NMR with CE–UV data are rather easier than the combination of data from other techniques (e.g. LC–MS and GC–MS). The difference between the two analytical techniques in terms of the contribution to the overall study may be explained by the different capacity of each analytical technique to visualize different types of compounds within a biological sample. For example, for this study oxalate and nitrate are observable in the CE–UV, but not in the ^{1}H NMR spectra, whereas taurine and the methylamines are more readily observable in the NMR profiles. In addition, in the multivariate models resulting from this study, NMR showed more changes between genders, while CE showed more changes in the type (knockout vs wild type). However, both techniques contributed information on different metabolites to the total metabolic profile, and extended the coverage of the metabolome. Moreover, the metabolites that are detected by both techniques provide confirmation of the models, and thus the two analytical technologies are highly complementary.

The combination of two analytical platforms (especially with NMR which provides structural information) allows for the application of statistical spectroscopy tools for compound identification (e.g. biomarker discovery), and relative relationships (e.g. metabolic connections and pathways). Specifically for NMR–CE, a program developed in Matlab based on the Pearson correlation coefficient, known as *capillary electrophoresis−nuclear magnetic resonance statistical hetero–analytical correlation spectroscopy* (CE−NMR–SHY) was applied in order to establish and plot the single linear correlations among variables deriving from the two different techniques. These correlations were driven from selected identified metabolite signals in the NMR or CE profiles that differentiated the slc26a6–null mice from the wild type[15]. In this chapter we showed the direct relationship between urinary oxalate (detected by CE–UV) and a gut macrobiotic metabolite (m–hydroxyphenylpropionyl sulfate, m–HPPS) along with a range of urinary metabolites from NMR–CE. In addition m–HPPS was tentatively identified in the CE profile; the confirmation could not be finished as the reagent standard is not available. The first work combining data from NMR and CE–UV platforms was done investigating *Schistosoma mansoni* infection in a mouse model[15]. Although this chapter is not the first work, this proves the advantages and the high value of this approach in the metabolomics field. However, it has not been proven that the combination of data from techniques such as NMR and CE could make the assignation of already established biomarkers in NMR in a profile in CE–UV possible.

REFERENCES.

1. Barbas, C.; Moraes, E. P.; Villaseñor, A., Capillary electrophoresis as a metabolomics tool for non–targeted fingerprinting of biological samples. *J Pharm Biomed Anal* **2011,***55* (4), 823–31.

2. Manz, A.; Pamme, N.; Iossifidis, D., *Bioanalytical chemistry*. Imperial College Press: 2004.

3. Ramautar, R.; Somsen, G.; de Jong, G., CE–MS in metabolomics. *Electrophoresis* **2009,***30* (1), 276–91.

4. Soga, T., Capillary electrophoresis–mass spectrometry for metabolomics. *Methods Mol Biol* **2007,***358*, 129–37.

5. Ramautar, R.; Demirci, A.; de Jong, G. J., Capillary electrophoresis in metabolomics. *Trends in Analytical Chemistry* **2006,***25* (5), 455–466.

6. García–Pérez, I.; Vallejo, M.; García, A.; Legido–Quigley, C.; Barbas, C., Metabolic fingerprinting with capillary electrophoresis. *J Chromatogr A* **2008,***1204* (2), 130–9.

7. Ahmed, F. E., The role of capillary electrophoresis–mass spectrometry to proteome analysis and biomarker discovery. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,***877* (22), 1963–81.

8. Barbas, C.; Vallejo, M.; García, A.; Barlow, D.; Hanna–Brown, M., Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies. *J Pharm Biomed Anal* **2008,***47* (2), 388–98.

9. Monton, M. R.; Soga, T., Metabolome analysis by capillary electrophoresis–mass spectrometry. *J Chromatogr A* **2007,***1168* (1–2), 237–46; discussion 236.

10. García, A.; Barbas, C.; Aguilar, R.; Castro, M., Capillary electrophoresis for rapid profiling of organic acidurias. *Clin Chem* **1998,***44* (9), 1905–11.

11. García–Pérez, I.; Whitfield, P.; Bartlett, A.; Angulo, S.; Legido–Quigley, C.; Hanna–Brown, M.; Barbas, C., Metabolic fingerprinting of Schistosoma mansoni infection in mice urine with capillary electrophoresis. *Electrophoresis* **2008,***29* (15), 3201–6.

12. Balderas, C.; Villaseñor, A.; García, A.; Rupérez, F.; Ibañez, E.; Señorans, J.; Guerrero–Fernández, J.; González–Casado, I.; Gracia–Bouthelier, R.; Barbas, C., Metabolomic approach to the nutraceutical effect of rosemary extract plus omega–3 PUFAs in diabetic children with capillary electrophoresis. *J Pharm Biomed Anal* **2010,***53* (5), 1298–304.

13. Guillo, C.; Barlow, D.; Perrett, D.; Hanna–Brown, M., Micellar electrokinetic capillary chromatography and data alignment analysis: a new tool in urine profiling. *J Chromatogr A* **2004,***1027* (1–2), 203–12.

14. Garcia-Perez, I.; Villasenor, A.; Wijeyesekera, A.; Posma, J. M.; Jiang, Z.; Stamler, J.; Aronson, P.; Unwin, R.; Barbas, C.; Elliott, P.; Nicholson, J.; Holmes, E., Urinary metabolic phenotyping the slc26a6 (chloride-oxalate exchanger) null mouse model. *J Proteome Res* **2012,** *11* (9), 4425-35.

15. Garcia-Perez, I.; Couto Alves, A.; Angulo, S.; Li, J. V.; Utzinger, J.; Ebbels, T. M.; Legido-Quigley, C.; Nicholson, J. K.; Holmes, E.; Barbas, C., Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating Schistosoma mansoni infection in a mouse model. *Anal Chem* **2010,** *82* (1), 203-10.

Journal of proteome research

Article

pubs.acs.org/jpr

Urinary Metabolic Phenotyping the slc26a6 (Chloride−Oxalate Exchanger) Null Mouse Model

Isabel Garcia-Perez,[⊥,†,§,▽] Alma Villaseñor,[⊥,‡] Anisha Wijeyesekera,[†,▽] Joram M. Posma,[†,§,▽] Zhirong Jiang,[¶] Jeremiah Stamler,[∥] Peter Aronson,[¶] Robert Unwin,[#] Coral Barbas,[‡] Paul Elliott,[§,▽] Jeremy Nicholson,[⊥,†,▽] and Elaine Holmes*[,⊥,†,▽]

[†]Biomolecular Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, SW7 2AZ, London, U.K.

[‡]CEMBIO (Center for Metabolomics and Bioanalysis) Pharmacy Faculty, Campus Monteprincipe, Universidad CEU San Pablo, Madrid, Spain

[§]Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, U.K.

[∥]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States

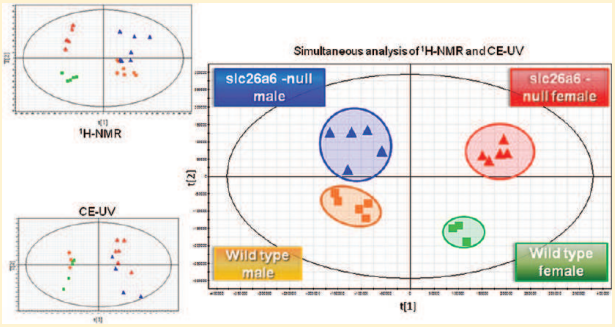
[¶]Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, United States

[#]UCL Centre for Nephrology, Royal Free Hospital, University College London, London, U.K.

[▽]MRC-HPA Centre for Environment and Health, Imperial College London, London, U.K.

S *Supporting Information*

ABSTRACT: The prevalence of renal stone disease is increasing, although it remains higher in men than in women when matched for age. While still somewhat controversial, several studies have reported an association between renal stone disease and hypertension, but this may be confounded by a shared link with obesity. However, independent of obesity, hyperoxaluria has been shown to be associated with hypertension in stone-formers, and the most common type of renal stone is composed of calcium oxalate. The chloride−oxalate exchanger slc26a6 (also known as CFEX or PAT-1), located in the renal proximal tubule, was originally thought to have an important role in sodium homeostasis and thereby blood pressure control, but it has recently been shown to have a key function in oxalate balance by mediating oxalate secretion in the gut. We have applied two orthogonal analytical platforms (NMR spectroscopy and capillary electrophoresis with UV detection) in parallel to characterize the urinary metabolic signatures related to the loss of the renal chloride−oxalate exchanger in slc26a6 null mice. Clear metabolic differentiation between the urinary profiles of the slc26a6 null and the wild type mice were observed using both methods, with the combination of NMR and CE-UV providing extensive coverage of the urinary metabolome. Key discriminating metabolites included oxalate, *m*-hydroxyphenylpropionylsulfate (*m*-HPPS), trimethylamine-*N*-oxide, glycolate and *scyllo*-inositol (higher in slc26a6 null mice) and hippurate, taurine, trimethylamine, and citrate (lower in slc26a6 null mice). In addition to the reduced efficiency of anion transport, several of these metabolites (hippurate, *m*-HPPS, methylamines) reflect alteration in gut microbial cometabolic activities. Gender-related metabotypes were also observed in both wild type and slc26a6 null groups. Urinary metabolites that showed a sex-specific pattern included trimethylamine, trimethylamine-*N*-oxide, citrate, spermidine, guanidinoacetate, and 2-oxoisocaproate. The gender-dependent metabolic expression of the consequences of slc26a6 deletion might have relevance to the difference in prevalence of renal stone formation in men and women. The different composition of microbial metabolites in the slc26a6 null mice is consistent with the fact that the slc26a6 transporter is found in a range of tissues, including the kidney and intestine, and provides further evidence for the "long reach" of the microbiota in physiological and pathological processes.



KEYWORDS: *oxalate, slc26a6, CFEX, NMR spectroscopy, metabolic profiling, capillary electrophoresis*

INTRODUCTION

There is an epidemiological association between renal stone formation and other disorders such as hypertension,[1,2] type 2 diabetes, and obesity. However, the mechanistic and aeitopathogenic links between these apparently diverse conditions are not fully understood.[1] The most common stone type is composed of calcium oxalate, and for this reason, there has always been interest in oxalate metabolism and excretion in renal stone disease. An animal model that has shed light on the importance of altered oxalate balance in renal stone formation is the slc26a6 (CFEX, PAT-1) null mouse, originally produced because of

Received: December 22, 2011
Published: May 17, 2012

the expression of this chloride−oxalate (formate) exchanger in the renal proximal tubule, and a postulated role in renal sodium reabsorption and thus blood pressure control.[3,4] Moreover, a large-scale metabolome-wide association study found an inverse correlation between urinary formate levels and diastolic blood pressure.[5] However, to date there is no clear evidence for a direct link between the function of this exchanger and blood pressure. Indeed, slc26a6 has a broad specificity, and in addition to oxalate and formate, it can also transport chloride, sulfate and bicarbonate; transport is also inhibited by *p*-aminohippurate.[7] This transporter is not found exclusively in the kidney, but it is also present in the gut (small intestine, cecum, and proximal colon), pancreas, skeletal muscle, and heart.[3] However, its highest affinity is for oxalate, which has been the main focus of recent attention, prompted by the serendipitous finding of calcium oxalate renal stone formation in slc26a6 null mice due to significant hyperoxaluria secondary to reduced gut secretion of oxalate, leading to increased circulating levels.[3]

A dual metabolic profiling strategy, combining NMR spectroscopy with CE-UV of the same urine sample, has been applied in order to comprehensively describe the urinary metabolic phenotype of the slc26a6-null mouse model. Both technologies have been widely applied in the metabolic fingerprinting of various diseases, including hypertension, type 2 diabetes, and inflammatory conditions, and together offer complementary information. While NMR generates comprehensive and robust profiles of a broad range of low molecular weight molecules, CE-UV can provide quantitative data on ionic species including chloride and nitrates. Here we show the direct relationship between urinary oxalate and a range of urinary metabolites and show clear sex-dependent slc26a6-related differences in metabolism.

MATERIALS AND METHODS

Sample Collection

Urine samples were provided by the Department of Internal Medicine, Yale University School of Medicine. Wild type control and slc26a6 null mice were bred continuously on a 129S6/SvEv genetic background. The urine samples were obtained from the mice at 6 weeks old. Protocols for use of experimental animals were approved by the Yale Institutional Animal Care and Use Committee. Animal handling and generation of slc26a6 null mice is detailed elsewhere.[4]

Chemicals

Sodium tetraborate decahydrate (STD), β-cyclodextrin sulfated (analytical grade, Sβ-CD) and methanol were purchased from Sigma-Aldrich (Steinheim, Germany), sodium dodecyl sulfate (SDS) and sodium hydroxide from Panreac Química S.A.U. (Barcelona, Spain), and hydrochloric acid from Fluka (Buchs, Switzerland). Reverse osmosed deionized water (Milli-Q Synthesis from Millipore, Bedford, MA, USA) was used for standard solution and electrolyte preparations. All the standards used for peak identification were obtained from Sigma Aldrich.

^{1}H NMR Sample Treatment and Instrument Conditions

Urine samples were prepared with pH 7.4 phosphate buffer as described previously.[8] ^{1}H NMR spectroscopy was performed at 300 K on a Bruker 800 MHz spectrometer (Bruker Biospin, Karlsruhe, Germany) using the following standard one-dimensional pulse sequence with saturation of the water resonance: Relaxation delay (RD)-90°-t_1-90°-t_m-90°-acquire free induction decay (FID), where 90° represents the applied 90° radio frequency (rf) pulse, t_1 is an interpulse delay set to a fixed interval of 3 ms, RD was 2 s and t_m (mixing time) was 100 ms. Water suppression was achieved through irradiation of the water signal during RD and t_m. Each urine spectrum was acquired using 8 dummy scans, 256 scans, 64 000 time domain points with a spectral width of 20 000 Hz. Prior to Fourier transformation, the FIDs were multiplied by an exponential function corresponding to a line broadening of 0.3 Hz.

CE-UV Sample Treatment and Instrumental Conditions

Metabolic fingerprinting was performed using two different separation modes:[9,10] a capillary zone electrophoresis method in reverse polarity (CZE-RP) that detects mainly anions, and a micellar electrokinetic chromatography method in normal polarity (MEKC-NP) that detects cations and neutral compounds.

CE experiments were carried out on a P/ACE MDQ system (Beckman Instruments, Fullerton, CA, USA) equipped with diode array UV-absorbance detection (190−600 nm), a temperature-controlled (liquid cooled) capillary compartment and an autosampler. Electrophoretic data were acquired and analyzed with 32 Karat software (P/ACE MDQ instrument).

Separations achieved with reverse polarity were performed using a capillary (Beckman Coulter, Buckinghamshire, England) coated with polyacrylamide (PAG) 60 cm in total length, 50 cm effective length and 50 μm internal diameter. On first use, the capillary was conditioned by a pressure flush of 0.1 M HCl (1 min), electrophoretic buffer (BGE1) (10 min) and an electrokinetic flush of electrolyte with 0.5 kV/cm (10 min). Between runs, the capillary was flushed under pressure with deionized water (2 min) and BGE1 (2 min). All experiments were performed at 25 °C using a separation potential of −20 kV. Samples were injected at the cathode, with 0.5 psi (3447 Pa) pressure applied for 10 s. The compounds were detected at the anode. BGE1 was prepared with 0.2 M phosphoric acid, adjusted to pH 6.10 with NaOH, and 10% (v/v) methanol added (the current observed under these conditions was 70 μA). Data were collected at a frequency of 4 Hz.

Separations achieved with normal polarity were performed in a fused silica capillary, 50 cm total, 40 cm effective length (50 μm internal diameter) (Composite Metal Services, Hallow, Worcester, U.K.). New capillaries were conditioned for 30 min at 25 °C with 1 M NaOH, followed by 0.1 M NaOH for 20 min and deionized water for 10 min. Before each analysis, the capillaries used were washed with 0.1 M HCl and deionized water for 1 min, and then 2 min with the run buffer. The running buffer (BGE2) comprised 25 mM sodium borate, 75 mM SDS and 6.25 mM sulfated β-cyclodextrin. The pH was adjusted to pH 9.50 with 1 M NaOH, after addition of SDS and cyclodextrin. Buffer solutions were filtered through a 45 μm filter before use. The capillary was maintained at 20 °C, with 18 kV applied voltage and 5 s hydrodynamic injection.

All the samples were initially run in CZE-RP method using an acidified urine sample. Afterwards, the same samples were mixed with 5 μL of BGE2 and 40 μL of water for MEKC-NP method.

Peak identification was the main drawback of CE-UV due to the selectivity of the UV detector. We were also limited by the small amount of sample. Peak assignment was performed by spiking the sample with standards and comparing migration times and spectral absorption.

Preprocessing and Data Analysis

^{1}H NMR Spectroscopy. The spectra were automatically phased, baseline corrected and referenced to 3-(trimethylsilyl)

dx.doi.org/10.1021/pr2012544 | J. Proteome Res. 2012, 11, 4425−4435

propionate-2,2,3,3-d_4 (TSP; δ 0.0) using an in-house routine written in MATLAB (The MathWorks, Natick, Massachusetts). Spectral regions corresponding to water and urea resonances (δ 4.5−6.5) were removed, and spectra were normalized using the probabilistic quotient method.[11]

The NMR peak assignments were based on published literature chemical shifts.[12,13] Chemical shifts belonging to the same metabolite were assigned by statistical total correlation spectroscopy (STOCSY) using an in-house Matlab script.[14]

CE-UV. Raw profiles from both polarities were treated with an in-house program developed in Matlab 7.0. First they were baseline corrected and aligned separately using the correlation optimized warping (COW) method previously described;[15] later both polarities were joined, first the signals from MEKC-NP mode followed by signals from CZE-RP profile. The maximum value for hippurate, a signal that appears in both profiles, was obtained, and the ratio between both hippurate signals was used for scaling. Subsequently all the data were normalized to the probabilistic quotient,[11] and the resultant matrix was submitted to multivariate data analysis with SIMCA P+ 12.0.

NMR-CE. Raw data from NMR and CE were combined in a single matrix after baseline correction, normalization and multialignment had been independently achieved. Both data sets were scaled to the highest peak in the profile to give equal weight to all signals.

Principal component analysis (PCA) was performed with the combined data set (^{1}H NMR, CE-UV and NMR-CE) in order to generate metabolic profiling data. Moreover, partial least square discriminant analysis (PLS-DA) and orthogonal partial least square discriminant analysis (OPLS-DA)[16] were also carried out to characterize the metabolic consequences of slc26a6 deletion in mice.

Every model was built using pareto scaling,[17] and the quality of the models was assessed by the cumulative R^2 value, and the predictive ability by cumulative Q^2 extracted according to the internal 7-fold cross-validation[18] default method of Umetrics SIMCA-P+12.0.1 software. The significance of the discriminatory metabolites was validated using a combination of cross validation and jackknife intervals.

Furthermore for NMR-CE, a program developed in Matlab based on the Pearson correlation coefficient, known as capillary electrophoresis−nuclear magnetic resonance statistical heteroanalytical correlation spectroscopy (CE−NMR-SHY), was applied in order to establish and plot single linear correlations among variables deriving from the two different techniques.[19] These correlations were driven from selected identified metabolite signals in the NMR or CE profiles that differentiated the slc26a6-null mice from the wild type. They were further analyzed to aid peak identification and to establish possible pathway associations with other metabolites. As in previous studies,[19] correlations greater than a threshold θ of 0.3 for this CE−NMR-SHY, corresponding to high sensitivity and a high probability of detecting structural correlations, were plotted.

RESULTS AND DISCUSSION

Metabolic Phenotyping of the slc26a6 Null Mouse Model

Clear metabolic differentiation was found between slc26a6 null mice and wild type control mice for both analytical techniques, namely NMR spectroscopy (Figure 1A,B) and CE-UV (Figure 2A,B). The CE-UV data presented are the result of the combination of two modes of CE separation in which the total profile contains cations, anions and neutral compounds.[20,21] A strong metabolic signature discriminating male and female mice was also apparent in both data sets. Several of the strain dependent metabolic differences were found to correlate with urinary oxalate.

Characterization of the Metabolic Phenotype of slc26a6 Null Mice Using ^{1}H NMR Spectroscopy. Significant changes were observed in the global ^{1}H NMR metabolic fingerprint of the SLC26A6-null mice urine, showing the strong impact of genetic modulation on the urine composition (Figure 1A,B). These differences manifest in the PCA scores plot as driving the variation in the second component (t2) (Figure 3A). The first component (t1) was dominated by variation relating to sex differences specifically: trimethylamine (TMA 2.88 (s)) and spermine (3.12 (m), 2.138 (m)) were higher in concentration in male mice and glycerate (4.09 (dd)), formate (8.46 (s)) and citrate (2.68 (d), 2.54 (d)) were higher in female mice. Regardless of sex differences, the slc26a6-null phenotype was dominated by higher urinary concentrations of (*m*-HPPS) (7.38 (t), 7.21 (m), 2.91 (t), 2.51 (t), 7.16 (dd)) and lower urinary hippurate signals (7.84 (d), 7.64 (t), 7.56 (t), 3.97 (d)) in comparison with the wild type mice (Figure 1D). The PCA model coefficients also indicated higher urinary levels of trimethylamine-*N*-oxide (TMAO) and creatine were associated with slc26a6 null mice with respect to the wild type mice.

OPLS-DA models were built to compensate for the effect of gender on the model in order to obtain a more direct profile of the slc26a6-related influence. In addition to the differential slc26a6 related metabolites identified from the PCA models, OPLS-DA also ranked scyllo-inositol, lactate and alanine as significantly higher in the urine of slc26a6 mice, whereas urinary excretion of hippurate, taurine and TMA was lower in the slc26a6 group.

Characterization of the Metabolic Phenotype of slc26a6 Null Mice Using CE-UV. The CE-UV derived metabolic phenotype of slc26a6 null mice was stronger than the NMR derived phenotype (Figure 2A,B), with separation of slc26a6 null and wild type control mice dominating the first component of the PCA model and sex differences influencing the second component (Figure 3B). Most metabolites were structurally identified, but some significantly discriminatory signals could not be assigned to known metabolites since the identity of metabolites from CE-UV analysis is reliant on comparison of migration time and spectral absorbance of the pure standard; this is only achieved by careful spiking of samples. As can be seen from Figure 2C, slc26a6 null mice showed higher levels of *m*-HPPS (Var_6718), oxalate (Var_5727) and allantoin (Var_2122). On the other hand, hippurate levels (Var_2220 in normal polarity and Var_9726 in reverse polarity) were consistently higher in wild type control mice.

The difference between the two analytical techniques in terms of the contribution of genetic background and gender to the overall variance sources of variance can be attributed to the differential capacity of each analytical technique to visualize different types of compounds within a biological sample. For example, oxalate and nitrate are observable in the CE-UV but not the ^{1}H NMR spectra, whereas taurine and the methylamines are more readily observable in the NMR profiles. However, both techniques contribute information on different metabolites to the total metabolic profile and together extend the coverage of the metabolome. Moreover, the metabolites that are detected by both techniques provide confirmation of the models, and thus the two analytical technologies are highly complementary.

The main discriminatory features of slc26a6 deletion in mice in the CE-UV PCA model were higher concentrations of *m*-HPPS and oxalate and lower concentrations of hippurate in comparison

dx.doi.org/10.1021/pr2012544 | J. Proteome Res. 2012, 11, 4425−4435

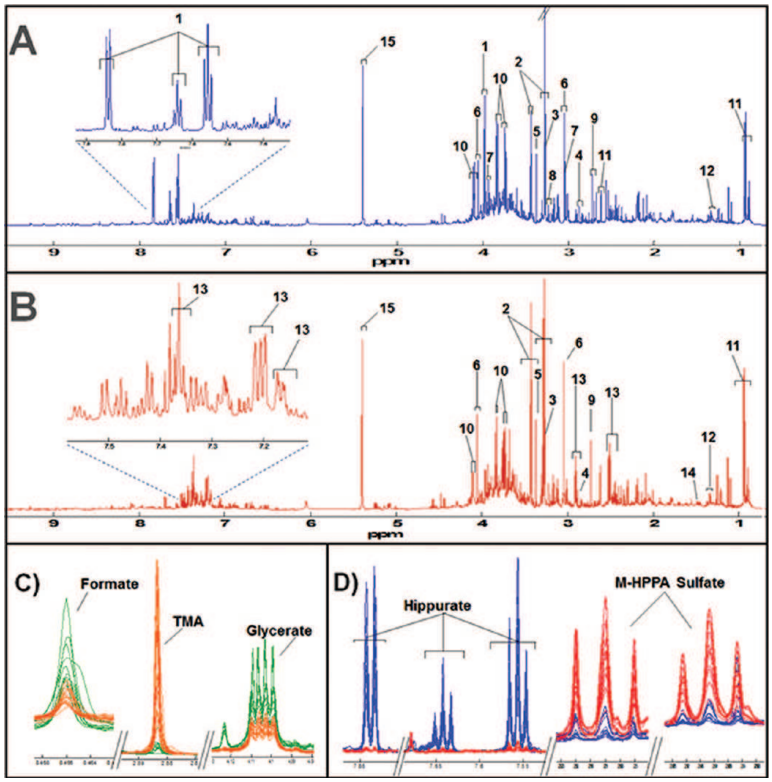


Figure 1. Typical 800 MHz ^{1}HNMR spectra of urine obtained from a male wild type mouse (A) and a male null-SLC26A6 mouse (B). Fragments of the NMR spectra (data aligned and normalized) showing differences in some of the discriminant metabolites according to the gender (C) and knockout condition (D). 1. Hippurate [δ 7.84 (d), 7.64 (t), 7.56 (t), 3.97 (d)]; 2. Taurine [δ 3.43 (t), 3.27 (t)]; 3. Trimthylamine *N*-oxide (TMAO) [δ 3.27 (s)]; 4. Trimethylamine (TMA) [δ 2.88 (s)]; 5. Scyllo inositol [δ 3.36 (s)]; 6. Creatinine [δ 4.05 (s), 3.045 (s)]; 7. Creatine [δ 3.93 (s), 3.04 (s)]; 8. Glycerophosphoryl choline [δ 3.23 (s)]; 9. Dimethylamine [δ 2.72 (s)]; 10. Glycerate [δ 4.09 (dd), 3.83 (dd), 3.72 (dd)]; 11. α-Ketoisocaproate [δ 2.62 (d), 0.94 (d)]; 12. Lactate [δ 4.11(q), 1.33 (d)]; 13. *m*-HPPA sulfate (*m*-HPPS) [δ 7.38 (t), 7.21 (m), 7.16 (dd), 2.91 (t), 2.51 (t)]; 14. Alanine [δ 3.79 (qt), 1.48 (d)].

with the wild type group. The OPLS-DA model was able to identify further discriminatory features including higher excretion of allantoin and chlorides, whereas citrate was positively associated with the wild type control.

Characterization of the Metabolic Phenotype of slc26a6 Null Mice Combining Two Analytical Platforms: ^{1}H NMR Spectroscopy and CE-UV. Statistical integration of nuclear magnetic resonance (NMR) spectroscopy and capillary electrophoresis (CE) data has been previously demonstrated in a mouse model based on urinary metabolite profiles[19] with a view to combining their different strengths and advantages, with regard to the ability to differentiate between different biological conditions and to identify potential metabolic pathways of response.

This data integration strategy was applied to the metabolic characterization of slc26a null mice compared to wild type control using a combination of high resolution ^{1}H NMR spectroscopy and CE with ultraviolet detection (UV) in order to achieve a more complete global metabolic profile in which oxalate is included.

PCA score plots derived from the combined data set (Figure 3C) showed grouping of samples according to gender along the first component (t1) as well as clear separation between wild type control and slc26a6 null mice along the second component (t2), indicating that the metabolic characteristics of the groups are markedly different. The combination of both analytical techniques allowed visualization of a broader range of metabolites in the same profile. Moreover, the enhanced quality of the model built using the combined data set over the PCA models generated from individual analytical methods can be measured by the increase in predictive capability of the models (Q^2), which increased from 0.34 (CE) and 0.39 (NMR) to 0.45 for the combined NMR-CE (Figure 3).

The OPLS-DA model, used as before to reduce the influence of gender on the model, found clear separation between the slc26a6 null and wild type groups along the predictive component on the OPLS-DA cross-validated scores plots (Figure 4A) of the NMR-CE combined data set. Variables with higher covariance on the predictive OPLS-DA component were identified from the S-plots. Higher levels of *m*-HPPS, oxalate, scyllo-inositol, TMAO, creatine, lactate, alanine, chlorides, allantoin and glycolate were characteristic of slc26a6 null mice, while higher urinary levels of hippurate, taurine, TMA, succinate, methionine, citrate were features of wild type mice. The significance of these metabolites as characteristic of the genetic strain was confirmed using the jack-knife interval (Figure 4B, Supporting Information Table S1). When the interval includes zero value, the covariance is not significant and the compound should not be considered as a potential biomarker. For this reason, creatinine, although apparently discriminatory in the OPLS-DA models was excluded from the Supporting Information Table S1, which summarizes the potential candidate biomarkers indicated by the plot and lists the discriminant metabolites obtained individually by each technique. Spearman rank correlation between variables intensity and

dx.doi.org/10.1021/pr2012544 | *J. Proteome Res.* 2012, 11, 4425–4435

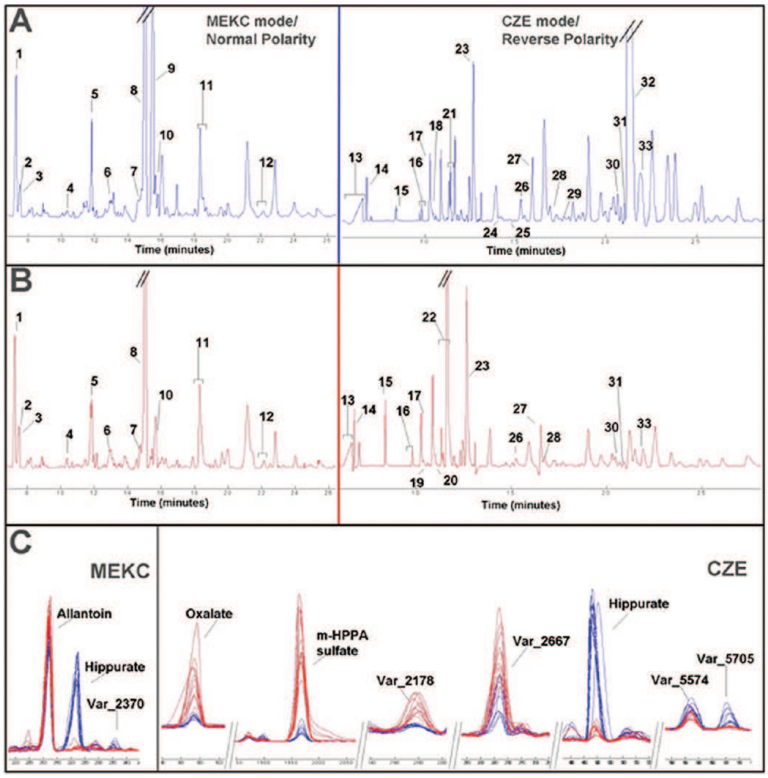


Figure 2. CE-UV metabolic fingerprinting of urine corresponding to male wild type mouse (A) and male SLC26A6-null mouse (B). Different levels observed in some metabolites according to the knockout condition (C). 1. Urea; 2. Creatinine; 3. Creatine; 4. Histidine; 5. Phenylalanine; 6. Methyluridine; 7. PAG; 8. Allantoin; 9. Hippurate; 10. Uridine; 11. Uric/Inosine; 12. Benzoate; 13. Chlorides; 14. Nitrate; 15. Oxalate; 16. Fumarate; 17. Oxoglutarate; 18. Succinate; 19. Isocitrate; 20. Glutarate; 21. Citrate; 22. *m*-HPPA sulfate (theorical); 23. Glycolate (coelutes); 24. Acetoacetate; 25. Lactate; 26. Ketosiovalerate; 27. Glycerate/Benzoate; 28. α-Ketoisocaproate; 29. Phenylpyruvate; 30. Phenyllactate; 31. Glutamate; 32. Hippurate/Vanillate 33. Ascorbate. Wild type control highlighted in blue. SLC26A6-null mice highlighted in red.

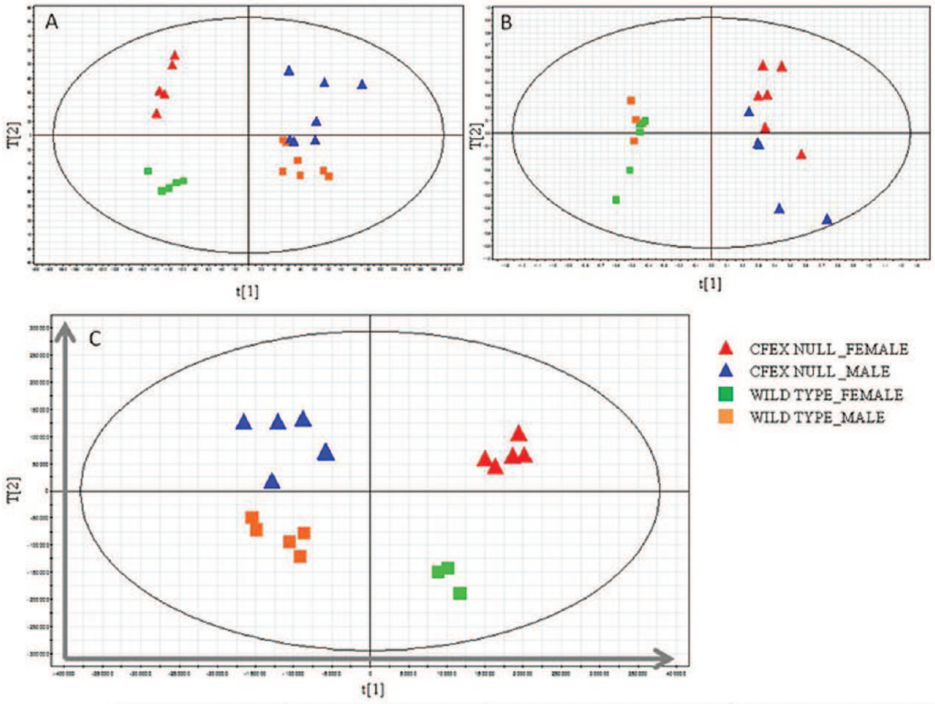
group for metabolites found to be significant using OPLSDA and jack-knifing is also provided in Supporting Information Table S1. The discrete distribution of the classes, slc26a6 null (n = 11) and wild type (n = 8), also causes a discrete distribution in the rank correlation. In this study, the highest absolute rank correlation (fully separated groups) that can be achieved is 0.856 with a corresponding p-value of 2.89×10^{-6}, because of the discrete group levels indicating a true and marked difference in the metabolism of the SLC26A6 null mice. Finally, using the p-values determined with the Spearman correlation, we have calculated the corresponding q-values using the false discovery rate[22] (FDR) using the R-package fdrtool.[23] We chose to allow 5% false discoveries ($q \leq 0.05$) for our findings. The metabolites that remain positively associated with slc26a6 deletion, after FDR, are *m*-HPPA, lactate, glycerolphosphorylcholine and oxalate. Hippurate remains negatively associated with slc26a6 deletion.

Influence of Gender on the Metabolic Phenotype of slc26a6 Null Mice. On the basis of the results obtained from the PCA model, OPLS-DA models were performed in order to study the influence of gender and to determine if the metabolic features of slc26a6 mice were consistent for male and female animals. Clear clustering in both the PCA and the OPLS-DA scores plots due to gender was also observed. Discriminant metabolites for female mice were identified by NMR and CE separately. Higher urinary concentrations of TMAO, creatine, 2-oxoisocaproate, 3-methyl-2-oxovalerate, *m*-HPPS, citrate and glycerate were associated with female mice in the NMR profiles. Conversely, TMA, methionine, spermine and alanine were positively associated with male mice. Further gender dimorphic metabolites extracted from the more OPLS-DA models of the CE-UV data were uridine (higher in females) and urea, penylalanine and allantoin (higher in males). As before, the statistical significance of the metabolites was confirmed by jack-knife interval (Figure 5A,B).

Supporting Information Table S2 shows the Spearman rank correlation (RHO correlation) between male (n = 11) and female (n = 8) mice. These result in the same discrete possible rank correlations as in the slc26a6 null versus wild type comparison in Supporting Information Table S1. Of the metabolites manifesting sex differences in urinary concentration in the current experiment, guanidinoacetate, creatine, creatinine, spermine, TMA and trimethylamine-*N*-oxide (TMAO) have previously been reported in mice.[1] Sex differences for urinary creatine and citrate concentrations have also previously been reported for humans with higher excretion in females,[2] consistent with the observations in the current study. Guanidinoacetate, also excreted in higher concentrations in female urine, is converted into creatine in the liver using *S*-adenosylmethionine as the methyl donor,[3] and therefore it is possible that this pathway is up-regulated in females.

Higher levels of TMAO have been reported in female Han Wistar rats than in male rats.[24] Hippurate, succinate, 2-oxoglutarate and dimethylglycine were elevated in the urine of females.[25] Of the metabolites strongly correlated with gender, creatine, uric acid, trimethylamine, methionine, alanine and TMAO were also correlated with the slc26a6 null group and thus deconvolution of the gender versus slc26a6 transporter deletion effects is complicated.

dx.doi.org/10.1021/pr2012544 | *J. Proteome Res.* 2012, 11, 4425–4435



PCA	NMR		CE		NMR-CE	
	R^2X(cum)	Q^2(cum)	R^2X(cum)	Q^2(cum)	R^2X(cum)	Q^2(cum)
1 Component	0.50	0.30	0.32	0.24	0.35	0.25
[1-2] Component	0.62	0.34	0.49	0.33	0.56	0.43
[1-2-3] Component	0.69	0.39	0.57	0.34	0.64	0.45

Figure 3. Scores plots from the PCA model derived from the NMR (A), CE (B), and NMR_CE combined data set (C) using pareto scaling (red triangle, knockout female; blue triangle, knockout male; green box, wild type female; orange box, wild type male).

Although the main source of TMA is the bacterial breakdown of choline in the intestine, this metabolite may have a wider role in physiology and may act as a signaling molecule. Methylamines, particularly TMA, have been shown to be present in higher concentrations in the urine of male mice and are lower in castrated mice.[26,27] Trace amine-associated receptors are thought to act as pheromone receptors in the nose,[28] a hypothesis that is supported by the fact that several of these trace amines such as TMA and isoamylamine can speed up puberty in female mice. Conversely, urinary concentrations of TMAO are higher in female mice. TMAO can be generated from the oxidation of TMA via hepatic flavin monooxygenase enzymes, predominantly FMO3,[27] which is derived from dietary constituents such as choline. However TMAO can also be directly absorbed from food. For example, cold water dwelling fish typically contain high levels of this metabolite, as it has antifreeze functions.

m-HPPS excretion has been shown both in the current study and previous studies to be influenced by gender. In the present study, it is also highly correlated with the slc26a6-null group. Stanley et al. reported higher concentrations of *m*-HPPS sulfate in male rat urine samples, which reflects the higher sulfation activity of male sulfotransferase enzymes in rats.[30] Sex differences in sulfation are well documented with four times higher cytosolic sulfation in male rat liver compared with females. However, this sex-related difference was reversed in mice, with female mice showing the higher sulfating activity, which fits with the higher levels of HPPS observed in female mice in the current study.[31]

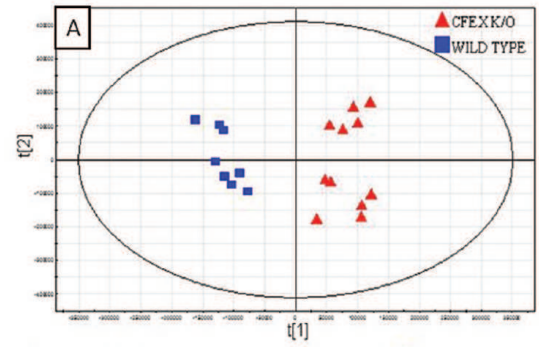
As for the slc26a6 deletion findings, we have calculated the corresponding *q*-values using FDR with respect to sex differences. For the gender comparison TMA, alanine, PAG and spermine remain positively associated with male mice after exclusion based on *q*-values. *m*-HPPA, formate, TMAO, creatine, creatinine, 2-oxoisocaproate, 3-methyl-2-oxoisovalerate, glycerate, guanidinoacetate and methylmalonate remain positively associated with female mice.[29]

Identification of Statistical Correlation between the Two Main Phenotypic Characteristics, Oxalate and *m*-HPPS with the Global Urine Profiles

Bidirectional correlation analysis of the combined data set (NMR-CE) was applied in order to establish pairwise correlations structures. It was also used for metabolic or functional correlations either within or between NMR and CE data,[19] such as those between molecules related by pathway or shared response to the slc26a6 deletion.

Since oxalate (Var_5727, *p*-value <0.05), a molecule that can be exchanged by slc26a6 and plays a key role in the characterization of slc26a6 deletion, is NMR invisible, a correlation matrix was calculated for the combined CE-UV and NMR detected signals using the oxalate signal in the CE-UV trace as the driver. A univariate Pearson correlation matrix encompassing all NMR and CE data was used with the oxalate signal with a cutoff level of $p < 0.05$. Signals that covaried across the NMR-CE data set are colored in red (Figure 6) and indicate a significant correlation ($p < 0.05$) with the oxalate signal. These signals included *m*-HPPS (signals on NMR; triplet δ 7.38 (r) 0.497, multiplet δ 7.21 (r) 0.559, triplet δ 2.91 (r) 0.561, triplet

dx.doi.org/10.1021/pr2012544 | *J. Proteome Res.* 2012, 11, 4425–4435



OPLSDA	NMR-CE CFEX KO vs WILD TYPE		
	R^2X (cum)	R^2Y (cum)	Q^2(cum)
1 Component	0.23	0.55	0.62
[1+2] Component	0.91	0.94	0.98

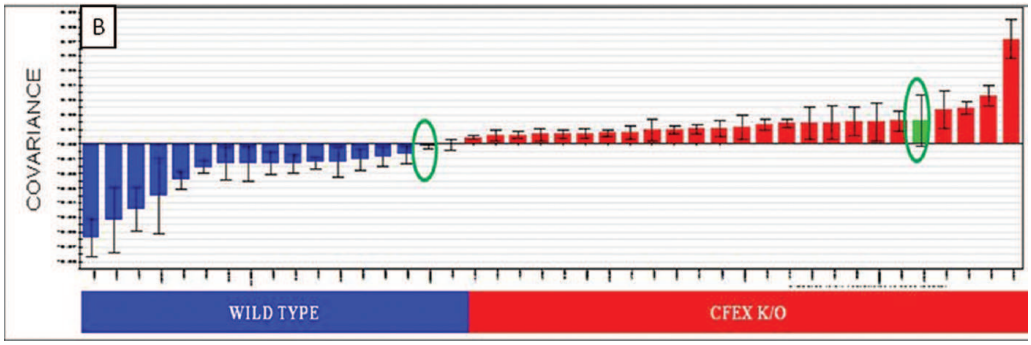
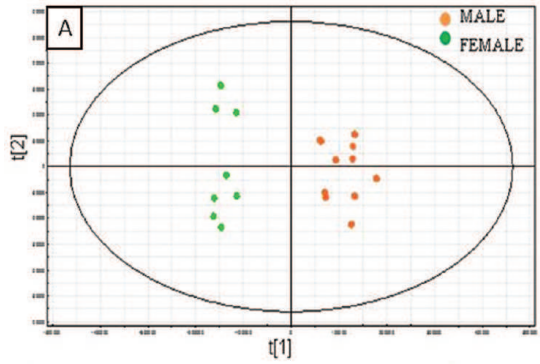


Figure 4. (A) Scores plot from OPLS-DA model derived from the NMR_CE data showing separation according to slc26a6-null and wild type mice. (B) Covariance for discriminant variables in the wild type and slc26a6-null mice including jack-knife interval (p > 95%); green bars show no statistical significance as determined by the interval which includes zero value.



OPLSDA	NMR-CE_ FEMALE VS MALE		
	R^2X (cum)	R^2Y (cum)	Q^2(cum)
1 Component	0.35	0.95	0.91
[1+2] Component	0.45	0.98	0.95

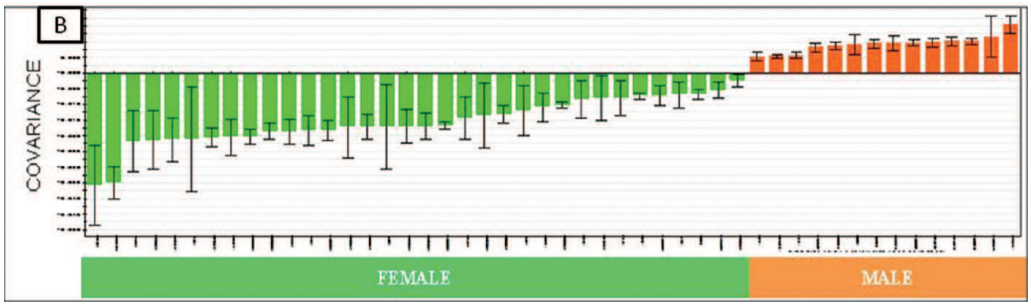


Figure 5. (A) Scores plot from the OPLS-DA model derived from the NMR_CE data showing separation according to gender (B) Covariance for discriminant variables in female and male mice including jack-knife interval (p > 95%).

δ 2.51 (r) 0.563, doublets of doublets δ 7.16 (r) 6.00, and in CE; Var_6718 (r) 0.886), TMAO (singlet δ 3.27 (r) 0.600), lactate (doublet δ1.33 (r) 0.610), chlorides (Var_5164 (r) 0.549), nitrate (Var_5228 (r) 0.718), fumarate (Var_6168 0.580), 2-oxoglutarate

dx.doi.org/10.1021/pr2012544 | J. Proteome Res. 2012, 11, 4425–4435

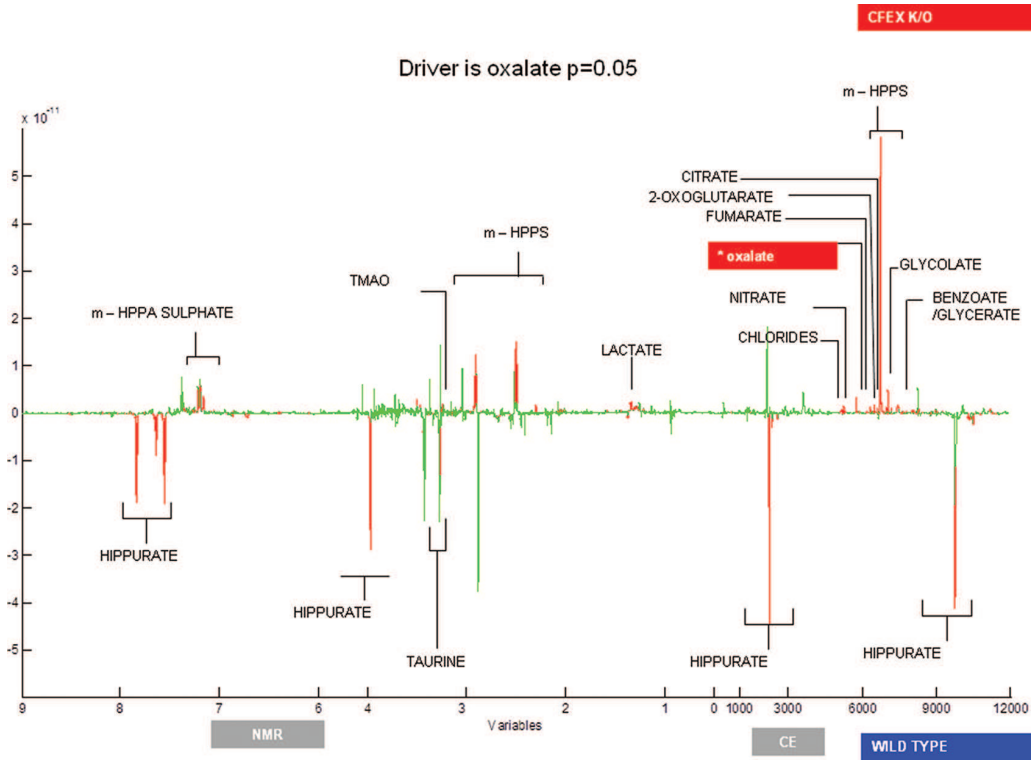


Figure 6. Correlation structure obtained after selecting oxalate in CE and applying a $p = 0.05$ significance criterion. Signals positively and highly correlated to oxalate signal can be seen highlighted in red on the top hemisphere of the figure, such as *m*-HPPS [signals on NMR; triplet δ 7.38 (r) 0.497, multiplet δ 7.21 (r) 0.559, triplet δ 2.91 (r) 0.561, triplet δ 2.51 (r) 0.563, doublets of doublets δ 7.16 (r) 6.00, and on CE; Var_6718 (r) 0.886], TMAO [singlet δ 3.27 (r) 0.600], lactate [doublet δ1.33 (r) 0.610], chlorides [Var_5164 (r) 0.549], nitrate [Var_5228 (r) 0.718], fumarate [Var_6168 0.580], 2-oxoglutarate [Var_6308 (r) 0.619], glutamate [Var_6528 (r) 0.730], citrate [Var_6628 (r) 0.757], glycolate [Var_7038 (r) 0.825], benzoate/glycerate [Var_8968 (r) 0.640]. Signals highly but negatively correlated to oxalate signal can be seen at the bottom hemisphere of the figure, also highlighted in red, hippurate [signals on NMR; doublet δ 7.84 (r) −0.599, triplet δ 7.64 (r) −0.595, triplet δ 7.56 (r) −0.596, doublet δ 3.97 (r) −0.600, and on CE; Var_2220 (r) −0.761, Var_9726 (r) −0.701], taurine [triplet δ 3.43 (r) −0.490, triplet δ 3.27 (r) −0.516].

(Var_6308 (r) 0.619), glutamate (Var_6528 (r) 0.730), citrate (Var_6628 (r) 0.757), glycolate (Var_7038 (r) 0.825), benzoate/glycerate (Var_8968 (r) 0.640), which also revealed positive correlation with oxalate. Moreover, hippurate (signals on NMR; doublet δ 7.84 (r) −0.599, triplet δ 7.64 (r) −0.595, triplet δ 7.56 (r) −0.596, doublet δ 3.97 (r) −0.600, and on CE; Var_2220 (r) −0.761, Var_9726 (r) −0.701), and taurine (triplet δ 3.43 (r) −0.490, triplet δ 3.27 (r) −0.516) were significantly anticorrelated with oxalate (Figure 6); *p*-values and correlation coefficients, corresponding to the signals for metabolites that statistically correlate with oxalate, can be seen in Supporting Information Table S3.

The general assumption would be that metabolites that are strongly significantly correlated are more likely to be related to each other either mechanistically, e.g., in a shared perturbed metabolic pathway, or undergoing a similar metabolic response to the genetic modification. Thus oxalate was most strongly associated with *m*-HPPS, glycolate, citrate, nitrate (all positive) and hippurate (negative). The association between oxalate and nitrate simply reflects the capacity of the transporter for carrying multiple anions.[8] Chloride was positively correlated to oxalate, but as it was added to every sample in order to ensure the complete oxalate solubility in the urine, its biological significance was not considered in the correlations. Glycolate is one of the precursors for oxalate, together with glycine, hydroxyproline and glyoxalate;[9] the other precursors were not indicated as discriminatory of the slc26a6 null model, nor were they significantly correlated with oxalate.

One of the strongest metabolic correlations with oxalate, and indeed one of the strongest metabolites characterizing the metabolic phenotype of the slc26a6 mouse, was *m*-HPPS. The correlation matrix driven from *m*-HPPS showed correlations with hippurate (negative) creatinine, methylmalonate, creatine and TMAO (positive). Here the picture is confounded slightly since *m*-HPPS, although highly discriminatory for the slc26a6-null class, is also consistently higher in female mice over males. Thus, some of the metabolites that were correlated to *m*-HPPS, such as creatinine and TMA, may be reflective of sex-related differences in metabolism rather than the genetic modification.

m-HPPS was found to be strongly discriminant for slc26a6 deletion in both the combined and gender stratified models, but was also modestly discriminant for gender. Since this metabolite was one of the most influential in characterizing the slc26a6 phenotype, a correlation network was also calculated in order to explore potential pathway connections.

A correlation matrix was calculated for the triplet δ 2.91, *p*-value = 1.93×10^{-5} of *m*-HPPS, as this was the signal with the highest significance in discriminating the slc26a6 null group and that also occurs in a spectral region with minimal spectral overlap from other metabolite resonances. As can be seen in Figure 7, TMAO (singlet δ 3.27 (r) 0.629), creatine (singlet δ 3.04 (r)

dx.doi.org/10.1021/pr2012544 | *J. Proteome Res.* 2012, 11, 4425–4435

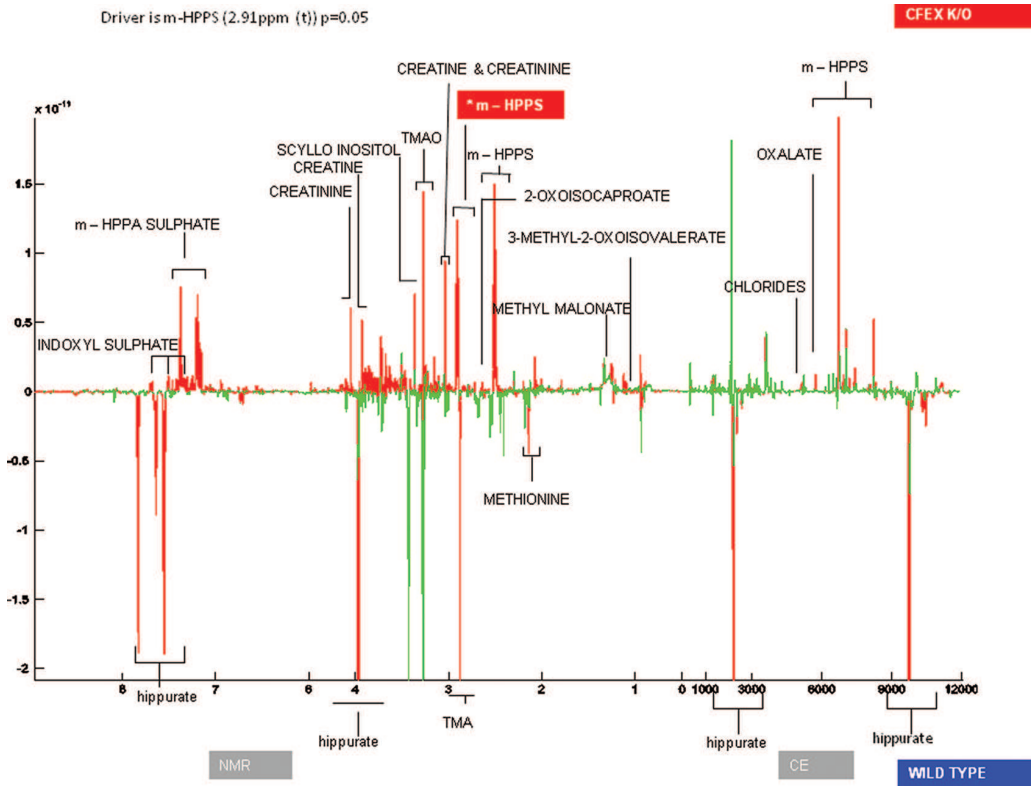


Figure 7. Correlation structure obtained after selecting *m*-HPPS on NMR and applying a $p = 0.05$ significance criterion. Signals positively and highly correlated to oxalate signal can be seen highlighted in red on the top hemisphere of the figure, such as TMAO [singlet δ 3.27 (r) 0.629], creatine [singlet δ 3.04 (r) 0.687, singlet δ 3.93 (r) 0.573], creatinine [singlet δ 3.045 (r) 0.738, singlet δ 4.05 (r) 0.753], scyllo-inositol [singlet δ 3.37 (r) 0.492], α-ketoisocaproate [doublet δ 0.94 (r) 0.535, doublet δ 2.62 (r) 0.655], methyl malonate [doublet δ 1.25 (r) 0.734], α-keto-β-methyl-*N*-valerate [doublet δ 1.11 (r) 0.573], indoxylsulfate [doublet δ 7.71 (r) 0.661, doublet δ 7.50 (r) 0.673, singlet δ 7.36 (r) 0.576], chlorides [Var_5164 (r) 0.178] and oxalate [Var_5727 (r) 0.483]. Signals highly but negatively correlated to oxalate signal, such as hippurate [signals on NMR; doublet δ 7.84 (r) −0.693, triplet δ 7.64 (r) −0.669, triplet δ 7.56 (r) −0.666, doublet δ 3.97 (r) −0.592, and on CE; Var_2220 (r) −0.739, Var_9726 (r) −0.760], TMA [singlet δ 2.88 (r) −0.642] and methionine [singlet δ2.14 (r) −0.564], can be seen at the bottom hemisphere of the figure, also highlighted in red.

0.687, singlet δ 3.93 (r) 0.573), creatinine (singlet δ 3.045 (r) 0.738, singlet δ 4.05 (r) 0.753), scyllo-inositol (singlet δ 3.37 (r) 0.492), 2-oxoisocaproate (doublet δ 0.94 (r) 0.535, doublet δ 2.62 (r) 0.655), methyl malonate (doublet δ 1.25 (r) 0.734), 3-methyl-2-oxovalerate (doublet δ 1.11 (r) 0.573), indoxylsulfate (doublet δ 7.71 (r) 0.661, doublet δ 7.50 (r) 0.673, singlet δ 7.36 (r) 0.576), chlorides (Var_5164 (r) 0.178) and oxalate (Var_5727 (r) 0.483), were positively and highly correlated to *m*-HPPS signal. In the same way, signals corresponding to hippurate (signals on NMR; doublet δ 7.84 (r) −0.693, triplet δ 7.64 (r) −0.669, triplet δ 7.56 (r) −0.666, doublet δ 3.97 (r) −0.592, and on CE; Var_2220 (r) −0.739, Var_9726 (r) −0.760), TMA (singlet δ 2.88 (r) −0.642) and methionine (singlet δ 2.14 (r) −0.564), were strongly negatively correlated to *m*-HPPS. Supporting Information Table S4 shows *p*-values and correlation coefficients of these metabolites.

The use of correlation matrices can provide additional information to that extracted from the multivariate models. For example, nitrate falls below the significance level set for the multivariate models generated for the CE-UV data but is highlighted as a correlate of oxalate using the Pearson correlation statistic. Likewise the correlation matrix calculated for *m*-HPPS identifies indoxylsulfate and several other gut microbial metabolites (TMA, TMAO, hippurate and other phenolics) as strong correlates, thereby further supporting the metabolic signature of gut microbial metabolites, although indoxyl sulfate is not in itself identified as a discriminatory feature of the OPLS-DA models.

Differential Expression of Gut Microbial Metabolites in the slc26a6 Null Mouse

The use of transgenic models of altered intestinal transport can also probe the interactions between dietary substrates and the gut microbiota. There is increasing interest in the gut microbiome and the symbiotic relationship with its host, more specifically the influence it may have on host immunity and metabolism.[13] However, an aspect that has not been considered yet in any detail is whether alterations in intestinal nutrient transport can in turn affect metabolism by changing the gut microbes or their metabolism. Intestinal bacteria have their own metabolic pathways and metabolites, which may integrate with those of the host.[13]

Although the slc26a6 urine model was created to explore the relationship between renal anion exchange and sodium regulation, it is known that this anion exchange transporter is ubiquitously found throughout the intestine and other tissues, such as heart, kidney, and liver.[3] The strong differences in *m*-HPPS and hippurate suggest that the metabolic phenotype of this knockout is related at least partially to altered intestinal microbial presence or activity, which may be more closely associated with effects arising from the expression of this transporter in the intestine. *m*-HPPS is a metabolite product from the metabolism of caffeic acid (CA), ferulic acid (FA) and chlorogenic acid,

dx.doi.org/10.1021/pr2012544 | *J. Proteome Res.* 2012, 11, 4425–4435

which appears to occur in the intestinal mucosa and colonic microflora.[4] However, the major pathway forming *m*-HPPA is the reduction of caffeic acid to 3,4-dihydroxyphenylpropionic acid followed by dehydroxylation to *m*-HPPS by the action of the gut flora. Aromatic amino acids, including tyrosine, tryptophan and phenylalanine, are also dietary precursors of phenolic acid-related metabolites. The metabolic end points of the intestinal metabolism of tyrosine include phenylpropionic acid and *m*-HPPS.[10] The rate of excretion of *m*- and *p*-hydroxyphenylproionic acid has been shown to be low in germ free animals, since most of the *m*-hydroxyphenylproionic acid involves reduction of caffeic acid followed by para-dehydroxylation, which does not occur in germ free rats. *m*-HPPA excretion increases dramatically (~140 fold) upon inoculation with fecal micro-organisms,[11] which further supports the microbial origin of this metabolite.

Although the urinary concentrations of hippurate and *m*-HPPS do not appear to be stoichometrically related, other than oxalate levels, they represent the two most discriminatory metabolites for the slc26a6 null model and are anticorrelated. Hippurate is formed from the metabolism of dietary aromatic compounds to benzoate by the gut microbiota and subsequent conjugation of benzoate with glycine in the mitochondria. The concentration of urinary hippurate has been shown to be modulated according the composition of the intestinal microbiota,[15] and rats have been shown to switch from excreting hippurate as the predominant microbial metabolite to *m*-HPPA and vice versa in as little as a few hours.[32]

As mentioned previously, a Pearson correlation calculated for a selected *m*-HPPS signal identified a range of microbial metabolites or microbial–mammalian cometabolites including indoxylsulfate, TMA, TMAO, and hippurate.

The significance of urinary TMA and TMAO has been discussed in relation to sex differences. Altered urinary excretion of TMAO has been observed in previous studies, most commonly in cases of drug-induced nephrotoxicity,[5] but this metabolite has also been positively associated with cardiovascular disease in both humans and mouse models.[33] Feeding of hyperlipidemic mice with either choline or TMAO was shown to result in higher plasma levels of TMAO, which was directly associated with the size of atherosclerotic plaque. Higher plasma TMAO concentrations were found in female mice, which also developed more severe atherosclerosis than male mice.[33]

Another indication of potential gut microbial involvement with oxalate excretion is that intestinal colonization with the bacterium Oxalobacter formigenes, a normal gut commensal, appears to promote intestinal oxalate secretion and reduce urinary oxalate excretion in rats and mice.[12,30] Thus, although the role of the gut microbiota in defining the slc26a6 null mouse model is far from clear, the results point to the fact that microbial metabolites are differentially expressed in this model of anion transport. Taken together with the literature on choline metabolism, there may be a link with the hypertension associated with renal stone formation, but more detailed laboratory experiments using labeled substrates would be required to probe this link further.

CONCLUSIONS

In summary, we have found a strong metabolic phenotype for slc26a6 deletion, which is defined both by altered urinary composition of anions such as oxalate, chloride and nitrates and also by modified urinary metabolites deriving from gut microbial metabolism including *m*-HPPS, hippurate and methylamines. Differential levels of tricarboxylic acid cycle intermediates were also contributory to the metabolic signature of the slc26a6 null mice. The modification of the microbial metabolites is consistent with the fact that the slc26a6 transporter is found in a range of tissues other than the kidney, including the intestine. Sex-related differences were also found in the metabolic phenotype of slc26a6 null mice, which may relate to sex differences in renal stone formation and further may have implications in hypertension.

ASSOCIATED CONTENT

Supporting Information

Table S1: Metabolites observed in significantly different concentrations in urine samples obtained from slc26a6 null mice compared with wild-type control mice. Table S2: The changes of metabolites observed in urine obtained from female mice compared with male mice. Table S3: Summary of the metabolites that are positively and negatively associated with Oxalate (variable 5727), based on Pearson correlation. Table S4: Summary of the metabolites that are positively and negatively associated with *m*-HPPS (2.51 (t) ppm), based on Pearson correlation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +44 (0) 20 7594 3220. Fax: +44 (0) 20 7594 3226. E-mail: elaine.holmes@imperial.ac.uk.

Author Contributions

⊥These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The INTERMAP Study has been supported by Grants R01-HL050490 and R01-HL084228 from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA and by the Chicago Health Research Foundation. CEMBIO authors acknowledge EADS-CASA. Alma Villaseñor thanks Santander Bank for the mobility grant. Paul Elliott acknowledges funding support from the National Institute for Health Research (NIHR) Biomedical Research Centre, Imperial College Healthcare NHS Trust. He is an NIHR senior investigator.

REFERENCES

(1) Worcester, E. M.; Coe, F. L. Nephrolithiasis. *Primary Care* **2008**, *35* (2), 369–91.

(2) Johri, N.; Cooper, B.; Robertson, W.; Choong, S.; Rickards, D.; Unwin, R. An update and practical guide to renal stone management. *Nephron Clin. Pract.* **2010**, *116* (3), c159–71.

(3) Aronson, P. Essential roles of CFEX-mediated Cl(-)-oxalate exchange in proximal tubule NaCl transport and prevention of urolithiasis. *Kidney Int.* **2006**, *70* (7), 1207–13.

(4) Jiang, Z.; Asplin, J.; Evan, A.; Rajendran, V.; Velazquez, H.; Nottoli, T.; Binder, H.; Aronson, P. Calcium oxalate urolithiasis in mice lacking anion transporter Slc26a6. *Nat. Genet.* **2006**, *38* (4), 474–8.

(5) Holmes, E.; Loo, R. L.; Stamler, J.; Bictash, M.; Yap, I. K.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I. J.; Veselkov, K. A.; Daviglus, M. L.; Kesteloot, H.; Ueshima, H.; Zhao, L.; Nicholson, J. K.; Elliott, P. Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **2008**, *453* (7193), 396–400.

(6) Parmar, M. S. Kidney stones. *BMJ [Br. Med. J.]* **2004**, *328* (7453), 1420–4.

(7) Aronson, P. Role of anion transporter SLC26A6 (CFEX) in prevention of hyperoxaluria and urolithiasis. *AIP Conf. Proc.* **2007**, *900* (1), 141–8.

(8) Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2007**, *2* (11), 2692−703.
(9) García, A.; Barbas, C.; Aguilar, R.; Castro, M. Capillary electrophoresis for rapid profiling of organic acidurias. *Clin. Chem.* **1998**, *44* (9), 1905−11.
(10) Vallejo, M.; Angulo, S.; García-Martínez, D.; García, A.; Barbas, C. New perspective of diabetes response to an antioxidant treatment through metabolic fingerprinting of urine by capillary electrophoresis. *J. Chromatogr., A* **2008**, *1187* (1−2), 267−74.
(11) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **2006**, *78* (13), 4281−90.
(12) Lindon, J.; Nicholson, J.; Everett, J. NMR Spectroscopy of biofluids. *Annu. Rep. NMR Spectrosc.* **1999**, *38*, 1−88.
(13) Martin, F. P.; Dumas, M. E.; Wang, Y.; Legido-Quigley, C.; Yap, I. K.; Tang, H.; Zirah, S.; Murphy, G. M.; Cloarec, O.; Lindon, J. C.; Sprenger, N.; Fay, L. B.; Kochhar, S.; van Bladeren, P.; Holmes, E.; Nicholson, J. K. A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. *Mol. Syst. Biol.* **2007**, *3*, 112.
(14) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem.* **2005**, *77* (5), 1282−9.
(15) Tomasi, G.; van den Bergand, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.* **2004**, *18*, 231−241.
(16) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J.; Holmes, E.; Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemom.* **2006**, *20*, 341−351.
(17) van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, *7*, 142.
(18) Eriksson, I.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis and Applications*; Umetrics: Umeå, Sweden, 2001; p 533.
(19) Garcia-Perez, I.; Couto Alves, A.; Angulo, S.; Li, J.; Utzinger, J.; Ebbels, T.; Legido-Quigley, C.; Nicholson, J.; Holmes, E.; Barbas, C. Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating *Schistosoma mansoni* infection in a mouse model. *Anal. Chem.* **2010**, *82* (1), 203−10.
(20) García-Pérez, I.; Whitfield, P.; Bartlett, A.; Angulo, S.; Legido-Quigley, C.; Hanna-Brown, M.; Barbas, C. Metabolic fingerprinting of *Schistosoma mansoni* infection in mice urine with capillary electrophoresis. *Electrophoresis* **2008**, *29* (15), 3201−6.
(21) Barbas, C.; Vallejo, M.; García, A.; Barlow, D.; Hanna-Brown, M. Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies. *J. Pharm. Biomed. Anal.* **2008**, *47* (2), 388−98.
(22) Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., Ser. B* **1995**, *57* (1), 12.
(23) Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **2008**, *24* (12), 1461−2.
(24) Stanley, E. G.; Bailey, N. J.; Bollard, M. E.; Haselden, J. N.; Waterfield, C. J.; Holmes, E.; Nicholson, J. K. Sexual dimorphism in urinary metabolite profiles of Han Wistar rats revealed by nuclear-magnetic-resonance-based metabonomics. *Anal. Biochem.* **2005**, *343* (2), 195−202.
(25) Williams, R. E.; Lenz, E. M.; Lowden, J. S.; Rantalainen, M.; Wilson, I. D. The metabonomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1H NMR and HPLC-TOF MS. *Mol. Biosyst.* **2005**, *1* (2), 166−75.
(26) Nishimura, N.; Zhang, J.; Abo, M.; Okubo, A.; Yamazaki, S. Application of capillary electrophoresis to the simultaneous determination of betaines in plants. *Anal. Sci.* **2001**, *17* (1), 103−6.
(27) Gavaghan McKee, C. L.; Wilson, I. D.; Nicholson, J. K. Metabolic phenotyping of nude and normal (Alpk:ApfCD, C57BL10J) mice. *J. Proteome Res.* **2006**, *5* (2), 378−84.
(28) Liberles, S. D. Trace amine-associated receptors are olfactory receptors in vertebrates. *Ann. N. Y. Acad. Sci.* **2009**, *1170*, 168−72.
(29) Dumas, M. E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C.; Mitchell, S. C.; Holmes, E.; McCarthy, M. I.; Scott, J.; Gauguier, D.; Nicholson, J. K. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (33), 12511−6.
(30) Hatch, M.; Gjymishka, A.; Salido, E. C.; Allison, M. J.; Freel, R. W. Enteric oxalate elimination is induced and oxalate is normalized in a mouse model of primary hyperoxaluria following intestinal colonization with Oxalobacter. *Am. J. Physiol.: Gastrointest. Liver Physiol.* **2011**, *300* (3), G461−9.
(31) Gong, D. W.; Murayama, N.; Yamazoe, Y.; Kato, R. Hepatic triiodothyronine sulfation and its regulation by growth hormone and triiodothyronine in rats. *J. Biochem.* **1992**, *112* (1), 112−6.
(32) Phipps, A.; Stewart, J.; Wright, B.; Wilson, I. Effect of diet on the urinary excretion of hippuric acid and other dietary-derived aromatics in rat. A complex interaction between diet, gut microflora and substrate specificity. *Xenobiotica* **1998**, *28* (5), 527−37.
(33) 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.; DiDonato, J. A.; Lusis, A. J.; Hazen, S. L. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **2011**, *472* (7341), 57−63.

Supplementary material

Supplementary table 1.Metabolites observed in significantly different concentrations in urine samples obtained from slc26a6 null mice compared with wild-type control mice.

Metabolite	Chemical Shift. ppm	NMR RHO	NMR pval	CE_Variable	CE RHO	CE pval
METABOLITES POSITIVELY ASSOCIATED WITH SLC26A6-null MICE						
m-hppasulphate **	7.38(t)	0.662	2.03E-03	Var_6718	0.817	1.93E-05
	7.21(m)	0.817	1.93E-05			
	2.91(t)	0.817	1.93E-05			
	2.51(t)	0.817	1.93E-05			
	7.16(dd)	0.817	1.93E-05			
Scyllo inositol	3.36(s)	0.428	6.74E-02			
TMAO **	3.27(s)	0.292	2.25E-01			
Creatine **	3.04(s)	0.448	5.46E-02	Var_3587	0.428	6.74E-02
	3.93(s)	0.214	3.79E-01			
Lactate	1.33(d)	0.701	8.34E-04			
Glycerolphosphorylcholine	3.23 (s)	0.817	1.93E-05			
Alanine *	1.48 (d)	0.214	3.79E-01			
Oxalate				Var_5727	0.701	8.34E-04
Chlorides				Var_5164	0.467	4.38E-02
Allantoin **				Var_2122	0.428	6.74E-02
Glycolate				Var_7036	0.37	1.19E-01
Uric/inosine *				Var_2910	0.214	3.79E-01
METABOLITES NEGATIVELY ASSOCIATED WITH SLC26A6-null MICE						
Hippurate	7.84(d)	-0.856	2.89E-06	Var_2220	-0.72	5.07E-04
	7.64(t)	-0.856	2.89E-06			
	7.56(t)	-0.856	2.89E-06	Var_9726	-0.701	8.34E-04
	3.97(d)	-0.856	2.89E-06			
Taurine	3.43(t)	-0.35	1.41E-01			
	3.27 (t)	-0.545	1.58E-02			
TMA *	2.88(s)	-0.428	6.74E-02			
Succinate	2.41(s)	-0.35	1.41E-01			
Methionine *	2.14(s)	-0.195	4.25E-01			
Citrate	2.68(d)	-0.331	1.66E-01	Var_6653	-0.291	2.25E-01
	2.54 (d)	-0.545	1.58E-02			

** Metabolites positively associated with male. ** Metabolites positively associated with female*

Supplementary table 2. The changes of metabolites observed in urines obtained from female mice compared with male mice.

Metabolite	Chemical Shift. ppm	NMR RHO	NMR pval	CE_Variable	CE RHO	CE pval
METABOLITES NEGATIVELY ASSOCIATED WITH FEMALE						
TMA	2.88(s)	0.856	2.89E-06			
Alanine	1.48(d)	0.778	8.64E-05			
Methionine	2.15(s)	0.058	8.12E-01			
Urea				Var_340	0.058	8.12E-01
Phenylalanine				Var_1338	0.058	8.12E-01
PAG	3.68(s)	0.856	2.89E-06	Var_2057	0.058	8.12E-01
	7.35 (m)	0.856	2.89E-06			
	7.43 (m)	0.856	2.89E-06			
Uridine				Var_2286	0.097	6.92E-01
Uric/ inosine				Var_2910	0.039	8.74E-01
Spermine	3.12 (m)	0.856	2.89E-06			
	2.13 (m)	0.856	2.89E-06			
METABOLITES POSITIVELY ASSCIATED WITH FEMALE						
m-hppasulphate	7.38(t)	0.778	8.64E-05	Var_6718	-0.039	8.74E-01
	7.21(m)	0.72	5.07E-04			
	2.91(t)	0.448	5.46E-02			
	2.51(t)	0.817	1.93E-05			
	7.16(dd)	0.545	1.58E-02			
Formate	8.46(s)	0.817	1.93E-05			
TMAO	3.27(s)	0.856	2.89E-06			
Creatine	3.04(s)	0.642	3.03E-03	Var_3587	0.545	1.58E-02
	3.93(s)	0.759	1.64E-04			
Creatinine	3.045(s)	0.856	2.89E-06	Var_3598	0.136	5.78E-01
	4.05(s)	0.856	2.89E-06			
2-oxoisocaproate	0.94(d)	0.856	2.89E-06			
3-methyl-	1.11(d)	0.117	6.34E-01			

2oxoisovalerate	0.90 (t)	0.856	2.89E-06			
Glycerate	3.83(dd)	0.856	2.89E-06			
Allantoin *				Var_2120	0.117	6.34E-01
Guanidino acetate	3.86(s)	0.856	2.89E-06			
Methyl malonate	1.25(d)	0.72	5.07E-04			

* Allantoin signals appear in the region, corresponding to water and urea resonances (δ 4.5-6.5), has been removed.

Supplementary table 3. Summary of the metabolites that are positively and negatively associated with Oxalate (variable 5727), based on Pearson correlation.

Metabolite	Chemical Shift. ppm	Correlation coeffcient *r*	pval	CE_Variable	Correlation coeffcient *r*	pval
m-hppasulphate	7.38(t)	0.497	3.03E-02	Var_6718	0.886	4.93 E -007
	7.21(m)	0.559	1.90E-02			
	2.91(t)	0.561	1.63E-02			
	2.51(t)	0.563	1.21E-02			
	7.16(dd)	0.6	3.23E-02			
TMAO	3.27(s)	0.6	5.61E-03			
Lactate	1.33(d)	0.61	5.86E-03			
Chlorides				Var_5164	0.549	2.46E-03
Nitrate				Var_5228	0.718	1.86E-03
Oxalate *				Var_5727 *	1	0.00E+00
Fumarate				Var_6168	0.58	4.07E-03
2-oxoglutarate				Var_6308	0.619	4.07E-03
Glutamate				Var_6528	0.73	1.06E-03
Citrate				Var_ 6628	0.757	3.00E-04
Glycolate				Var_7038	0.825	4.09E-05
Benzoate/ Glycerate				Var_8968	0.64	1.61E-02
Hippurate	7.84(d)	-0.599	-0.762	Var_2220	-0.762	0.000159
	7.64(t)	-0.595	-0.723			
	7.56(t)	-0.596	-0.701	Var_9726	-0.701	0.0175
	3.97(d)	-0.6	6.63E-03			
Taurine	3.43(t)	-0.49	4.40E-02			
	3.27 (t)	-0.516	1.97E-02			

* Variable selected for driving the correlation from.

Supplementary table 4. Summary of the metabolites that are positively and negatively associated with m-HPPS (2.51 (t) ppm) based on Pearson correlation.

Metabolite	Chemical Shift. ppm	Correlation coefficient *r*	pval	CE_Variable	Correlation coefficient *r*	pval
m-hppasulphate *	7.38(t)	0.959	7.36E-11	Var_6718	0.751	0.000000494
	7.21(m)	0.987	2.44E-15			
	2.91(t)	0.995	0.00E+00			
	2.51(t)*	1	0.00E+00			
	7.16(dd)	0.988	2.88E-15			
TMAO	3.27(s)	0.629	3.90E-03			
Creatine	3.04(s)	0.687	9.70E-03			
	3.93(s)	0.573	1.28E-02			
Creatinine	3.045(s)	0.738	3.10E-04			
	4.05(s)	0.753	2.00E-04			
Scyllo inositol	3.37(s)	0.492	1.98E-02			
α-Ketoisocaproate	0.94(d)	0.535	1.84E-02			
	2.62(d)	0.655	2.36E-03			
Methylmalonate	1.25(d)	0.734	3.40E-04			
α-Keto-β-methyl-N-valerate	1.11(d)	0.573	1.03E-02			
Indoxylsulphate	7.71(d)	0.661	2.00E-03			
	7.50(d)	0.673	1.58E-03			
	7.36(s)	0.576	6.01E-03			
Chlorides				Var_5164	0.178	2.28E-02
Oxalate				Var_5727	0.483	3.63E-02
Hippurate	7.84(d)	-0.693	4.50E-04	Var_2220	-0.739	4.60E-04
	7.64(t)	-0.669	1.74E-03			
	7.56(t)	-0.666	1.90E-03	Var_9726	-0.76	4.60E-04
	3.97(d)	-0.592	7.54E-03			
TMA	2.88 (s)	-0.649	2.62E-03			
Methionine	2.14(s)	-0.564	1.79E-02			

* Variable selected for driving the correlation from

CHAPTER 5

3. Liquid Chromatography–Mass Spectrometry (LC–MS)

3.1. Fundamentals of the technique

MS based metabolomics is a growing research area where large investments in instrumentation, resources and research efforts are being made. The past few years have seen dramatic advances in the technology; newer, more advanced instrumentation, reaching higher levels of analytical sensitivity, functionality and detection specificity. Liquid chromatography–mass spectrometry (LC–MS) is currently the most widely used MS technology in the bioanalytical sector[1-2]. This technique is very versatile, and can work in many forms because, while LC allows the use of different columns, mobile phases and gradients, additives and so on, MS not only works in positive and negative mode, but with different mass range, or scan modes (full or MS/MS mode). Furthermore, the instrumentation for LC and most MS can be very different, and so the characteristics, advantages and possibilities are as well. The number of metabolites measured will also depend on the resolution of the chromatographic system and the specificity of the detection technique. The general aspects of LC and MS techniques are discussed next.

3.2. Liquid chromatography basics

Chromatography (as mentioned in Chapter 2) is a physical separation method in which the components to be split are selectively distributed between two immiscible phases: a mobile phase which is flowing through a stationary phase bed[3]. When the mobile phase is a liquid, then the technique is called liquid chromatography (LC). A basic scheme of LC is shown in Figure 5.1. Here, the mobile phase is pumped through a column that contains the stationary phase. Then the sample is introduced to the column, which is the heart of the LC system, where the molecules can be separated through different mechanisms according to the different combinations of mobile and stationary phases. Finally, the components from the sample are detected and go to the waste.

From the different mechanisms of separation (Table 5.1) the most commonly used in metabolomics are the reverse phase methods, due to the robustness of the chromatographic results. In addition, RP–LC–MS have more reasons to be used; the metabolites that can be determined cover a broad range in molecular weights, polarities and physicochemical properties[4]. The variety of sorbent materials in the column is ample, but the most used sorbents in holistic approaches are C18 and C8 columns[5]. C18 columns are more suited to the analysis of selected amino acids, carnitines, bile acids and phospholipids, whereas C8 columns provide better separation and elution of lipids[6–7]. For RP–LC–MS, methanol or acetonitrile mixed with water, plus volatile additives such as formic or acetic acids or ammonium are the most commonly used mobile phases and additives. Nevertheless, RP–LC has a great disadvantage over the highly polar molecules and ionic analytes (which cover an important part of the metabolome) because they are poorly retained. In this case, hydrophilic interaction chromatography (HILIC) is a new column technology which offers a different selectivity, with better retention of polar analytes. HILIC conditions take place when a pseudostationary phase is formed with the support, the bonded phase and a water layer adsorbed onto the surface[4]. This mechanism of separation can offer not only an improved profile over the polar compounds, but with MS detection, a high sensitivity because of increased ionization efficiency resulting from the use of mobile phases containing a high proportion of organic solvents, (e.g., acetonitrile), which contribute to the generation of spray conditions important for achieving increased sensitivity[8]. A complete review on HILIC can be found elsewhere[4]. Nevertheless, HILIC still is not widely used in metabolic non–target analysis because of its poor reproducibility, although a research group has demonstrated the analysis of 200 urine samples using HILIC–UPLC[8]. For LC–MS analysis, conventional column formats are used, typically ranging from 2.1–4.6mm of i.d., 5–25cm in length and packed with 3–5µm particles[1].

With the development of instrumentation, the birth of ultra–high performance liquid chromatography (UHPLC) allows the use of high pressure level instruments (up to 15,000psi), enabling the use of columns with sub–2µm particle size and flow rates of up to 800µL/min[5, 8]. This results in large improvements in terms of analysis time (throughput), peak capacity, lower ionization suppression, signal–to–noise ratio and sensitivity[1, 4].

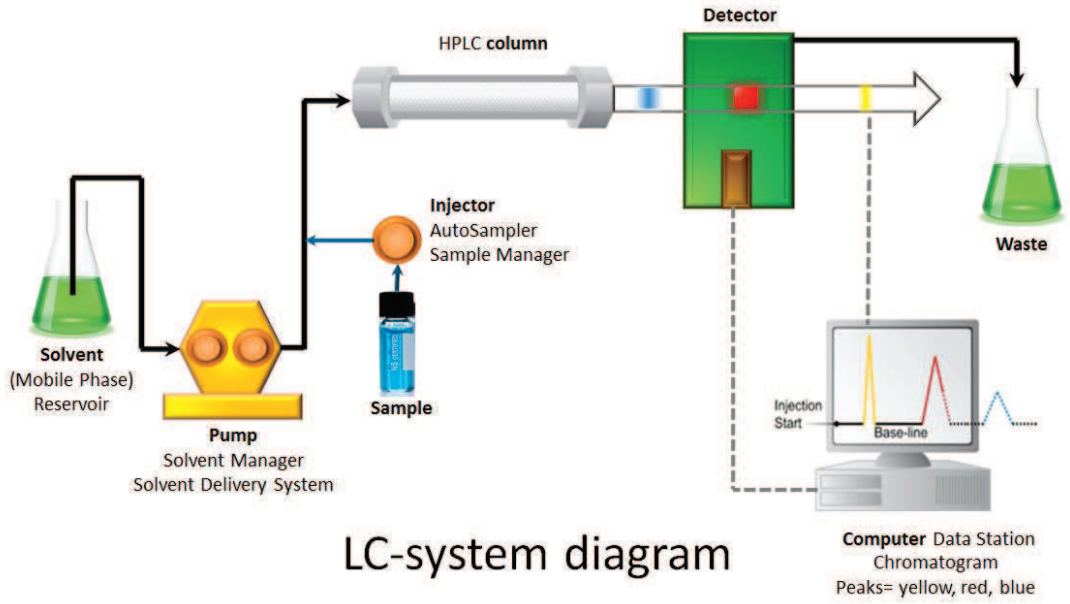


Figure 5.1. LC–system diagram adapted from waters webpage. http://www.waters.com/waters/en_US/.

Table 5.1 Separation mechanisms and phase systems in various LC modes (adapted from Wilfried M.A. Niessen, 2006[3]).

Separation mechanism	Characteristics	Mobile phase	Stationary phase	Solutes	Column for non–target analysis application
Adsorption (Normal phase)	Selective adsorption/desorption on a solid phase. Separation mechanism based on hydrophobicity and polarity	Non–polar organic solvent with organic modifier	Silica gel, alumina, bonded–phase material	Lipophilic: oils, fats, lipids	Done by aqueous normal phase (ANP). Column: Cogent diamond Hydride, 2.1mmX100mm, 4µm (MicroSolv Technology, Brisbane, Australia)[9–10]
Adsorption/Partition (Reversed phase)		Aqueous buffer with organic modifier (*e.g.*, MethOH, ACN)	Bonded–phase material, e.g., octadecyl–modified silica gel	Most of the bioorganic substances	Column: Discovery HS C18 15cmX2.1mm, 3µm; Supelco (Sigma Aldrich, St. Louis, MO,USA)[11]
Hydrophilic interaction (HILIC)	Involve different types of interactions: partition, ionic exchange and dipole–dipole interactions	Mixture of water plus organic (e.g., ACN)	Polar column (e.g., silica or amide bonded phase)	Highly polar compounds	Acquity BEH HILIC (1.7µm, 2.1 × 100mm) column (Waters Corporation, Milford, U.S.A.)[8]
Ion par	Formation of ion–par and selective partition or sorption of these ion–pairs	Aqueous buffer with organic modifier and pairing agent	Reversed–phase bonded phase material	Enantiomers	None
Partition	Selective partition between two immiscible liquids	Liquid, mostly non–polar	Liquid physically coated on porous solid support	Proteins, enzymes	None
Ion exchange	Differences in ion exchange properties	Aqueous buffers	Cationic or anionic exchange resin or bonded–phase material	Inorganic ions, acids, bases	None
Size exclusion	Differences in molecular size (more explicitly the ability to diffuse into and out of the pore system)	Non–polar solvent	Silica gel or polymeric material	Polymers, proteins, nucleic acids	None

3.3. Mass Spectrometry (MS)

The market of MS and LC–MS is extremely dynamic, and individual manufacturers are investing in the development of new technologies[12]. MS has been a standard technique used in the analytical research of molecules and complex mixtures for a long time. It is important in determining the elemental composition of a molecule and in gaining partial structural insights[13]. MS is based on the production of ions that are subsequently separated according to their mass–to–charge ratio (*m/z*). The resulting mass spectrum is a plot of the abundance of the generated ions (Figure 5.2).

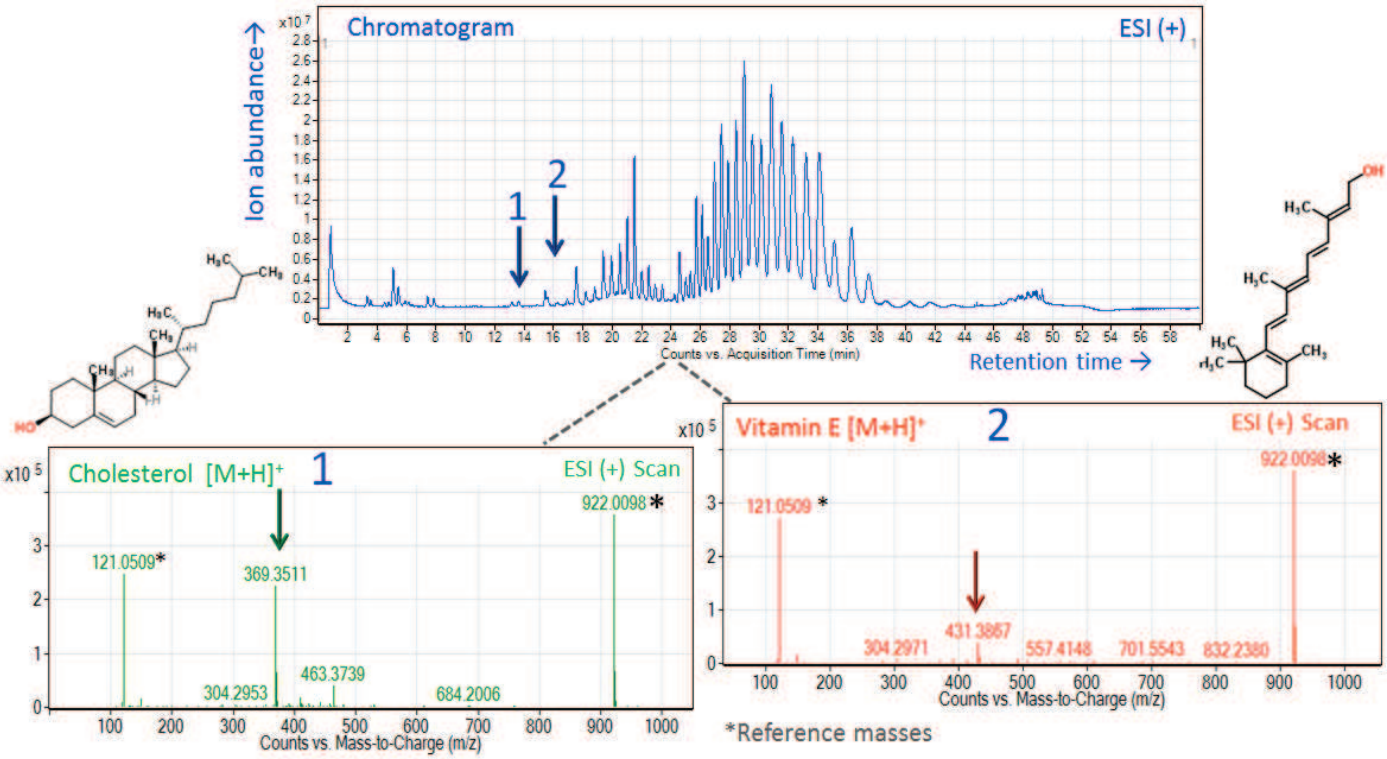


Figure 5.2. Total Ion Chromatogram in positive mode (TIC+) of a breast milk sample acquired by LC–QTOF–MS. Analytical conditions are described in chapter 2. Two peaks from the chromatogram have been taken showing the mass spectrum of cholesterol and vitamin E. Figure adapted from Khan, M. T. H., 2006[14].

Figure 5.2 shows the mass spectrum of common molecules, such as cholesterol and vitamin E, from a breast milk sample chromatogram obtained by LC–MS. Although the MS spectrometer is a highly sophisticated instrument, the process can be said to consist of: (*i*) sample introduction; (*ii*) ionization; (*iii*) mass analyser; (*iv*) ion detector; and (*v*) data handling (last two are out of the scope of this section).

(i) Sample introduction.

The sample introduction can be done directly through a continuous flux of sample or by interface with a separation technique such as LC. The direct infusion normally employs a syringe pump, a low velocity flux (< 20µL/min) and a sample volume of >10µL. This is frequently used for pure samples or a simple mix of compounds.

(ii) Ion source.

Ionization techniques are used to charge the molecules, because any compound that does not get charged will not reach the MS–detector. There are various analyte–ionization sources and they can primarily be classified as "*hard*" or "*soft*", depending on the extent of the fragmentation during the process. For *hard ionization*, the analyte is completely fragmented. This type of ion source is widely used in the interface coupling with GC technique, and thus was explained in depth in Chapter 2. For *soft ionization sources*, the eluted analyte from CE or LC, is ionized and transferred from liquid phase into gas phase. This is generally done by spraying the sample through a metal capillary surrounded by a turbulent air stream (e.g. N_2). Three major ionization methods are available: atmospheric pressure photo ionization (APPI); atmospheric pressure chemical ionization (APCI); and atmospheric pressure ionization (API), referred as electro spray ionization (ESI)[5, 12, 15]. There is a fourth ionization technique commonly used for macromolecules such as proteins and peptides, called matrix laser desorption/ionization (MALDI). From these, the method of choice in most metabolomics LC–MS applications is ESI[16].

The line out of the LC system is connected to a capillary, and the ESI ionization method relies on the formation of a spray or nebulization on exit of the capillary needle. ESI utilizes a high–electric field which in the positive mode of MS a positive voltage normally between 3,000 and 4,500V is applied to the capillary needle allowing the formation of charged analytes in solution. Here the charges are separated: the negative charges rapidly go to the wall of the capillary, while the positive are expelled, forming the

spray. Then this means that ions can be generated in solution. The vaporization of charged droplets from a liquid solution leads to coulombic explosions, which in turn release ions having multiple charge states *z* to the gas phase. The potential difference between the ESI capillary and the counter electrode ranges between 3,000V and 4,500V. Usually an electric field of 1000V/cm is maintained. The whole process is summarized in Figure 5.3. In metabolomics, the ESI source normally produces mainly ions with $z = 1$[5, 16].

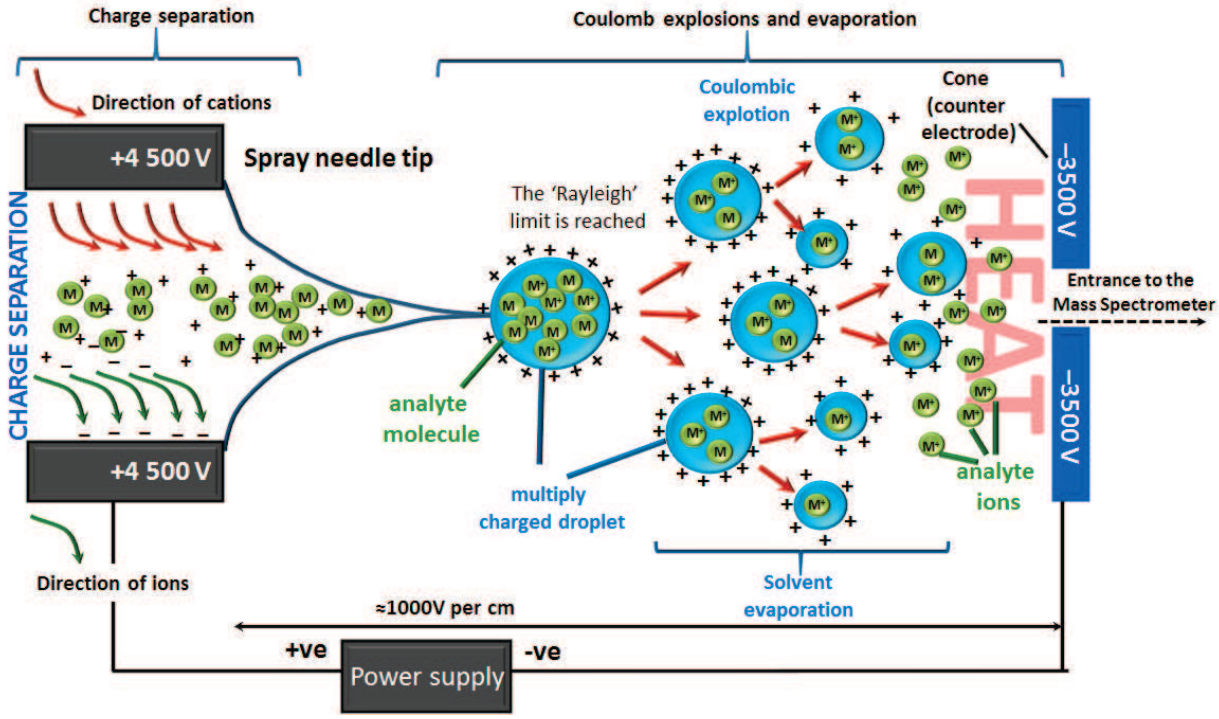


Figure 5.3. Scheme of the ESI ionization process adapted from Forcisi, S., 2012[5] and Bristol university web page: http://www.bris.ac.uk/nerclsmsf/techniques/hplcms.html.

Electrospray is the most efficient interface for injecting (normally producing $[M+H]^+/[M-H]^-$ ions) to mass analysers with little or no ion fragmentation (enabling the detection of a wide range of molecules). In addition, ESI could produce metallic adducts e.g. $[M+Na]^+$ or $[M+K]^+$ ions by thermal ion attachments. Cluster ions can be easily produced in the gas phase by non–covalent interactions between the produced ions and neutral solvent molecules, i.e. water. The formation of these species has the advantage that compounds unable to form clusters with hydrogen can be ionized on alternative routes such as sodium or potassium. As a result, detection of a wider metabolite range with ESI becomes possible. Benefits of this ionization technique also include excellent quantitative analysis, good (analyte–dependent) sensitivity, and ionization can be performed in positive and negative modes. The main advantages of the ESI ion source are soft ionization, no need for derivatization, ability to ionize compounds of a large mass range, suitability for non–volatile and from any polarity, excellent quantitative analysis and high sensitivity. Nevertheless, among the drawbacks of this is the probable statistical bias that could be observed during data analyses. Also, the ESI process is sensitive to changes in solvent composition as occurs in gradient chromatographic elution, causing poor vaporization/ ionization and unstable spray when solvents of high water content are used. So in order to stress the formation of different gas–phase cluster ions, additives should be injected into the mobile phase flow, such as formic acid, acetic acid, ammonium formiate, etc. Another inconvenience is the ion suppression. This is caused by the competition between the analyte and other components in the matrix, causing an increase or reduction of the analytical signal[7, 16]. For example, volatile salts and plasticizers may affect the size of the drops, the boiling temperature, and the surface tension of the liquid. In a sense, the LC system reduces the ion suppression compared to direct infusion in MS resolving co–eluting compounds, isomers and isobaric interferences in the case of low–resolving mass analysers. In addition, a good analytical separation will result in better detection limits and MS data quality, due to reduced background noise[17]. The ionization mechanism and application of the different interfaces are presented in Table 5.2.

Table 5.2. Characteristics of soft ionization techniques used in the interface in LC–MS[7, 16-17].

Soft ion source	Ionization mechanism	Applications
Atmospheric pressure ionization (API) → electrospray (ESI)	▪ Analytes are charged in solution and transferred to gas through the formation of a spray or by nebulization	▪ It is the softest of the ionization techniques, adequate for slight polar to ionic compounds. ▪ Ideal for thermally sensible compounds ▪ Dependent on the chemical solvent properties ▪ Sensibility is proportional to the concentration ▪ Allows the possibility to detect polymers of high MW due to the possibility of multiple charge
Atmospheric pressure chemical ionization (APCI)	▪ The nebulization gas is ionized through corona discharge ▪ The analytes are then charged by the collisions of the nebulization gas previously charged	▪ Valid for low to high polar compounds ▪ Analytes should have sort of volatility behaviour ▪ Good sensitivity for compounds of medium polarity and molecular weight
Atmospheric pressure photo ionization (APPI)	▪ Here the analyte is irradiated e.g. with krypton (Kr) light ▪ Ion–molecule reaction can be produced in the presence of protic solvents	▪ Suitable for very low and non–polar compounds ▪ Analytes do not need to be charged in solution ▪ Analytes should have sort of volatility behaviour
Matrix laser desorption/ ionization (MALDI)	▪ The sample is prepared by precipitation into a solid matrix that has basic/ acidic characteristics which through UV–vis radiations absorb energy causing the sublimation of the compound. During this desorption process, the charge is transfer from the matrix to the molecule	▪ Useful for macromolecules such as peptides, proteins, and oligosaccharides. ▪ The solid matrix protects the biomolecule from degradation making easier evaporation and ionization

(i) Mass analyser.

Once the ions enter the mass spectrometer, they are analysed. The purpose of the mass analyser is to measure the ions' mass–charge–ratio (*m/z*) or to select them. The basic common mass analysers include: single–quadrupoles(Q); triple–quadrupoles (QqQ); time–of–flight (TOF); ion traps (IT); orbitrap and fourier transform ion cyclotron resonance (FT–ICR). In addition, an increasing number of hybrid systems that combine two basic types of mass spectrometers are being introduced to the research field, such as quadrupole–time of flight QTOF, quadrupole linear ion traps (Q–Trap) or ion trap FT mass spectrometers[15]. The characteristics of the most widely used are shown in Table 5.3.

Table 5.3. Characteristics and common parameters of mass analysers used in LC–MS[12, 16, 18].

Mass Analyser	Characteristics	Resolving power [x10^3]	Mass accuracy (ppm)	Acquisition speed (Hz)	Tandem MS comments
Quadrupole (Q)	▪ Simplest option, robust, high–linear dynamic range (50–4,000 Da) ▪ Provides only nominal mass resolution (low resolution) Low sensibility to high masses ▪ Can work in scan or selected ion monitoring	▪ 3–5	▪ Low	▪ 2–10	▪ No applicable
Ion trap (IT)	▪ High sensitivity and acquisition speed ▪ Low resolution ▪ Possible to do tandem MS fragmentation ▪ Possibility of interactions ion–molecule	▪ 4–20	▪ Low	▪ 2–10	▪ MS^n, good accuracy and resolution ▪ Low energy collisions

Time of flight (TOF)	▪ Fast scanning capabilities, wide mass range, high resolution and mass accuracy ▪ Useful for profiling complex metabolic mixtures ▪ For MS/MS experiments TOF instruments has to be combined with another mass analyser, *i.e.* QTOF	▪ 10–60	▪ 1–5	▪ 10–50	▪ No applicable
Orbitrap and FT–ICR–MS	▪ Offer the highest resolution available and high–accuracy for fragments and masses (<2 ppm). ▪ FT–ICR–MS leads to a reliable formula candidates and supports metabolite identification	▪ 100–240	▪ 1–3	▪ 1–5	▪ MS^n, excellent accuracy and resolution of product ions
Triple quadrupole QqQ	▪ Tandem MS for structural investigations ▪ Capable of performing successive fragmentations ▪ Allows experiments such as neutral loss and precursor and product ion scan (multiple reaction monitoring, MRM and selected reaction monitoring SRM) ▪ Mainly applied to targeted metabolomics	▪ 3–5	▪ Low	▪ 5–15	▪ MS^2, good accuracy and resolution ▪ Low energy collisions
Quadrupole–time of flight QTOF	▪ This hybrid instrument combine the stability and robustness of the Q analyser with the fast scanning capabilities, accuracy and high sensitivity of TOF mass analysers	▪ 20–60	▪ 1–5	▪ 20–50	▪ MS^2, excellent accuracy and good resolution ▪ Low energy collisions

For this chapter, the mass analyser employed was quadrupole–time of flight (QTOF). For that reason, an in depth description is done next. Figure 5.4 illustrates a schematic representation of a QTOF mass analyser. When the charged ions enter, they pass directly to an ion guide (typically a quadrupole). Then they are transmitted through a pair of quadrupoles that (in scan mode) conducting the ions to the TOF part. Inside the TOF, the ions are being accelerated in an electric field reaching a terminal linear velocity which, depending on their *m/z* ratio, will vary over a constant distance (normally the tube is from 1 up to 2 meters long) from the acceleration point (pusher) to the detector. The TOF system involves measuring the time that ions take to travel from the beginning to the end of a field–free flight tube[5]. For low weight masses, their speed inside the TOF is higher, and so their time of flight is lower than high weight masses. Improving the resolving power and accuracy of this equipment, reflectrons or ion mirrors where introduced. They compensate for the flight time differences between ions of the exact same *m/z* ratio and increase the free ion flight path. In addition, an increase in accuracy is reached because during the analysis, reference masses are continuously measured, allowing constant mass correction.

A very important aspect for any MS system is the pressure. This should be low enough to ensure that there are few or no collisions of the ions with gas molecules or other ions, allowing free ion trajectories until the detector. Normally for TOF instruments, the pressure is approximately 10^{-7}Torr. The mass accuracy reached by the TOF is dependent on the constant length of the flight tube, based on the thermic isolation maintained at constant temperature, the high vacuum, the constant voltage in the ion pusher, and the use of reference masses for the calibration. The advantage of the high mass resolution is reflected not only in the accurate mass, but also in the isotopic distribution, which increases the possibility of a reliable compound identification due to lower possible chemical formulas for a particular mass.

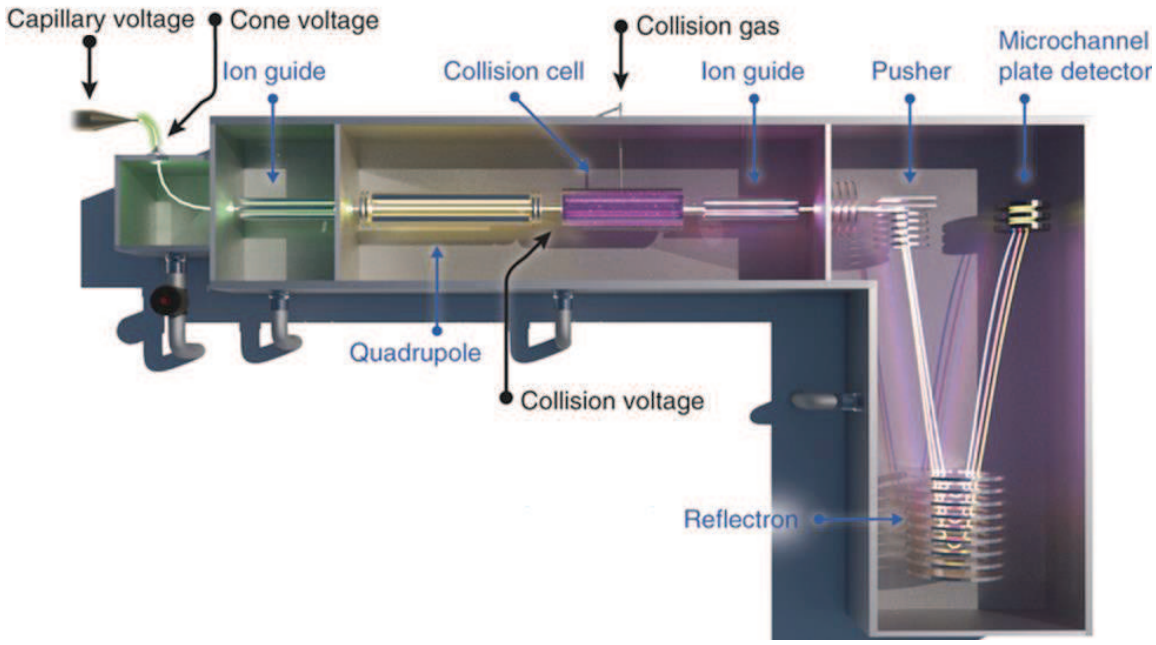


Figure 5.4. Schematic representation of quadrupole–time of flight (Q–TOF) mass spectrometer, adapted from[19].

Typically, the candidate biomarkers can be tentatively identified by consulting online databases such as METLIN, Lipid Maps, HMDB and KEGG, and the new online database mediator of CEU University, which covers all these databases except HMDB (http://ceumass.eps.uspceu.es/mediator/). These databases allow us to generate at a basic level a list of possible candidates based on monoistopic mass weight, within less than 10ppm of error, and possible chemical formula match from the isotope distribution. However, even with high mass accuracy instruments such as QTOF the list of candidates may still be long. This list could be edited with the help of some expertise. For example, hits of plant secondary metabolites, pharmaceutical, synthetic compounds or food additives could be discarded if they are irrelevant or out of the scope of the project. Also, when the data was acquired after e.g., reversed phase method using a C18 column, non–polar compounds are not expected within short retention time Therefore, the corresponding match in databases can be discarded. The article of Sandra, K., 2011[6] is a very good example of the characterization of metabolites expected by zones in the chromatogram in an LC–QTOF–MS for lipidomics approach in plasma. LC not only reduces mass spectral complexity, but also provides additional information on metabolite physicochemical properties, which help in metabolite

identification[1, 7, 17, 20]. All these aspects make it compulsory to confirm the tentative identification by MS/MS or through the use of an appropriate standard which are generally not commercially available. For the identification of metabolites in non–targeted metabolomic approaches, QTOF, orbitrap and Fourier transform ion cyclotron resonance mass spectrometers are instruments typically used[18].

MS/MS experiments increase confidence in a tentative identification by gleaning structural information from fragmentation data[17, 21]. The advent of hybrid MS, such as QTOF, allowed this integration by the use of the quadrupole. Therefore, QTOF instruments are frequently used for targeted and non–targeted approaches. In MS/MS mode, the biomarker candidate known as the precursor ion (expressed with a *m/z* ratio, RT, and mass/RT windows) will be selected and isolated by the first quadrupole (Q1) and then passed through to the collision cell (also known as collision induced dissociation (CID)) for fragmentation. The CID is a second quadrupole (q2) in which the energy is transferred to the selected ion precursor by collision with neutral molecules from a gas such as nitrogen, allowing the molecule to break up. The resulting ion fragments are then examined by the mass analyser TOF[22–23]. Here the fragments would be representative of the molecular structure of the precursor ion. The fragmentation patterns will be different for each instrument (IT, QqQ, QTOF)[24]. In the CID the energy applied can be set e.g., METLIN database contains a MS/MS fragmentation library of 10,787 compounds, usually recorded through MS/MS experiments at +/– 0, 10, 20 and 40Volts. The collision energy could be adjustable using a curve of mass weight versus energy. Here the energy will be different for each compound. For experimental work in this chapter, a slope of 3.6V/100Da for fragmentation was used. An example of the MS/MS fragmentation spectrum is shown in Figure 5.5. Here, the metabolite (*m/z*= 372.3109 and RT= 1.42 min) was fragmented in the CID using 175.0V. After the analysis, this was highly probably corresponding to tetradecanoylcarnitine.

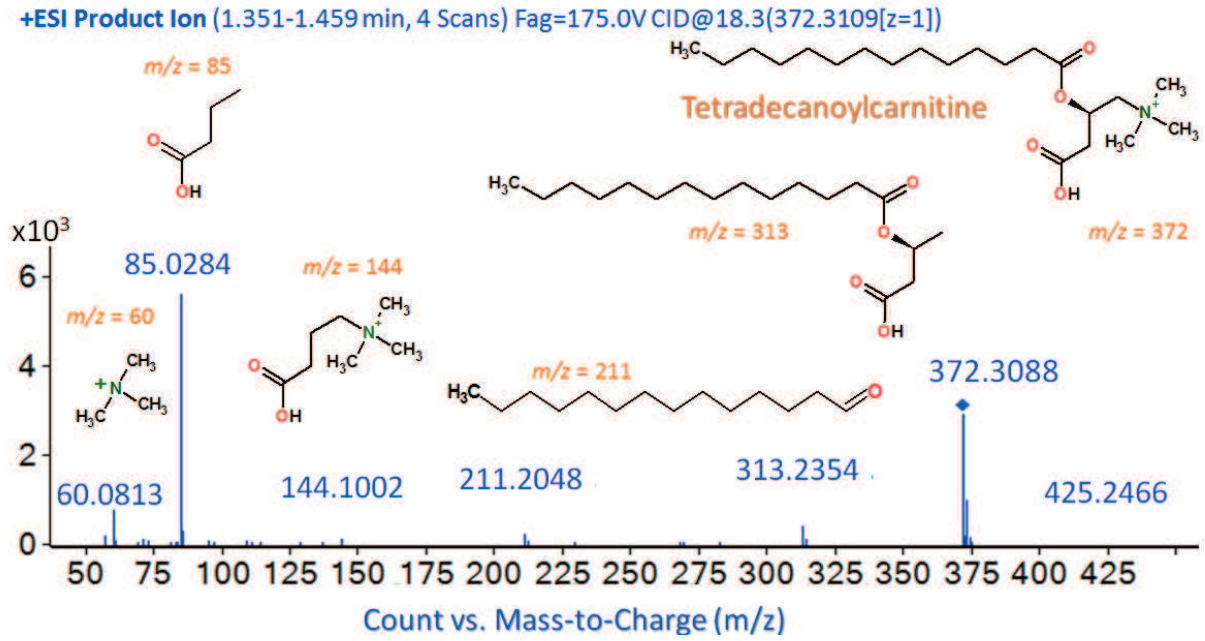


Figure 5.5. MS^2 spectra and structure of precursor ions for tetradecanoylcarnitine from brain tissue sample. NOTE: LC–MS conditions: 1µL of lipid extracted sample was injected into a RP column at 60°C (Agilent; Poroshell EC–C8, 15cm × 2.1mm, 2.7µm) with a pre–column (Supelco; Ascentis Express C8, 0.5cm× 2.1mm, 2.7µm). The system was operated in positive ion mode at a flow rate 0.5mL/min with solvent A (water with 5mM ammonium formate), and solvent B (methanol with 5mM ammonium formate and 10% of acetonitrile). The gradient started from 80% B to 95% B in 18min, then 100% B in 20min maintaining same gradient until 30min and returned to starting conditions in 1min, keeping the re–equilibration at 80%B for 14min. Data were collected in positive ESI mode in separate runs on a QTOF analyzer (Agilent 6520) operated in full scan mode from 50 to 1000 *m/z*. The capillary voltage was 3500V with a scan rate of 1.02 scans per second; the nebuliser gas flow rate was 10L/min

The interpretation of the fragments from MS/MS experiments is not an easy task. It can be hampered by the low concentration of the precursor ion, leading to a low resolution pattern. In addition, the introduction of other ions into the CID could cause misunderstanding of the results and it is possible that the precursor ion could be a hard molecule that may not have fragmented or been a fragile molecule causing lots of fragments, complicating the analysis. The fragmentation spectra can be compared with those at the database of METLIN as long as both were measured on the same MS instrument and with the same voltage energy. In the case that a different energy was used, the standard spectrum is only relative.

Nevertheless, it was observed that some typical groups of biomolecules followed similar fragmentation pattern, e.g., phospholipids (PC) and sphingomyelins (SM) contain choline so both show a characteristic *m/z* 184 fragment from the phosphocholine head group peak, and as well as a $[M+H–59]^+$ peak corresponding to the neutral loss of the trimethylamine $(CH_3)_3N$[25]. The fragmentation patterns for the most common classes of the lipid are shown in Table 5.4. Here the most common precursor ion and the typical fragments are mentioned.

Table 5.4. Lipid–specific mass spectrometry experiments. Precursor ion scan – PIS; neutral loss scan – NL [25–27].

	Biochemical category	Parent Ion	Fragments
POSITIVE MODE	CER	$[M+H]^+$	**PIS**, *m/z* **264** (sphingosine–H_2O)
	GL	$[M+H]^+$	**PIS**, *m/z* **264** (sphingosine–H_2O)
	SM	$[M+H]^+$	**PIS**, *m/z* **184** (phosphocholine), *m/z* **166** (phosphocholine–H_2O), *m/z* **104** (choline), *m/z***86** (choline–H_2O), NL of trimethylamine **$[M+H–59]^+$**
	TAG	$[M+NH_4]^+$	NL of fatty acids
	DAG	$[M+NH_4]^+$	NL of fatty acids
	CE	$[M+NH_4]^+$	**PIS**, *m/z* **369** (cholestane)
	PC & LPC	$[M+H]^+$	**PIS**, *m/z* **184** (phosphocholine), *m/z* **166** (phosphocholine–H_2O), *m/z* **104** (choline), *m/z* **86** (choline–H_2O), NL of water **$[M+H–H_2O]^+$**, for PC only: loss of one of the fatty acyl substituent as a ketene **$[M+H–LPC]^+$**
	PE & LPE	$[M+H]^+$	NL of phosphoethanolamine **$[M+H–141]^+$**
	PS & LPS	$[M+H]^+$	NL of phosphoserine **$[M+H–185]^+$**
	PG & LPG	$[M+NH_4]^+$	NL of glycerophosphate + NH_4 **$[M+NH_4–189]^+$**

NEGATIVE MODE	PE	$[M–H]^-$	**PIS**, *m/z* 196 (dilysophosphoethanolamine–H_2O), *m/z* **140** (phosphoethanolamine); NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**
	PS & LPS	$[M–H]^-$	NL of serine–H_2O **[M–H–87]$^-$**, NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**
	PG & LPG	$[M–H]^-$	**PIS**, *m/z* **153** (glycerophosphate–H_2O), *m/z* **171** (glycerol), NL of phosphatidic acid ion **[M–H–76]$^-$**, NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**
	PI & LPI	$[M–H]^-$	**PIS**, *m/z* **241** (inositolphosphate), *m/z* **223** (inositolphosphate–H2O), NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**
	PA & LPA	$[M–H]^-$	**PIS**, *m/z* **153** (glycerophosphate–H_2O), *m/z* **135** (glycerophosphate–$2H_2O$), NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**
	PC	$[M–Cl]^-$	**PIS** (either $[M+35]^-$ or $[M+37]^-$), NL of glycerophosphate–H_2O **[M–H–153]$^-$**, NL of phosphate (PO_4) **[M–H–97]$^-$**, NL of phosphite (PO_3) **[M–H–153]$^-$**

Abbreviations: Ceramides (**CER**); Glycolipids (**GL**); Triacylglycerides (**TAG**); Diacylglycerides (**DAG**); Cholesterol esters (**CE);** Phosphatidic acid (**PA**); Lysophosphatidic acid (**LPA**); Phosphatidylcholine (**PC**); lysophosphatidylcholine (**LPC**); phosphatidylethanolamine (**PE**); lysophosphatidylethanolamine (**LPE**); phosphatidylserine (**PS**); lysophosphatidylserine (**LPS**); phosphatidylinositol (**PI**); lysophosphatidylinositol (**LPI**); phosphatidylglycerol (**PG**); lysophosphatidylglycerol (**LPG**); sphingomyelin (**SM**).

As can be observed from Table 5.4, some compounds get charged only in negative mode, such as organic acids or carbohydrates[1], so in order to obtain a comprehensive profile, ionization should be performed in both positive and negative modes[16].

Normally the fragmentation pattern is not expected to explain all fragments, but will at least explain the most abundant (in my personal experience at least those fragments above 20% of relative abundance). In this difficult task certain software has been developed from commercial companies. Agilent, for example, has the Molecular Structure Correlator. This takes the raw data and compares the most abundant fragments obtained in the MS/MS spectra with the mass accuracy and the formula match score from the suggested molecule, giving a general score of relative reliability. The total confirmation of a compound ideally is by comparing the standard in the same analytical conditions, expecting close RT between the standard and the sample, and same fragmentation pattern[7].

LC–MS/MS has not only been used for identification; it has settled down as a quantitative target analysis approach. Among the different applications, is important to mention its role in clinical diagnosis, where typical metabolite screening is now a fact, e.g., about 30 parameters of amino acid and fatty acid metabolism are screened for the detection of inherited metabolic disorders[18].

3.4. Characteristics of LC–MS for metabolomics studies

All the particular characteristics of LC and MS have caused LC–MS to become increasingly popular for non–target analysis of biological samples, and in a wide variety of target applications in metabolomics research[28]. With advanced instrumentation now available, LC–MS–based metabolic profiling can provide sensitive, accurate and reproducible analysis of a wide range of metabolite classes present in biofluids, spanning a large dynamic range. With care, not only a detailed metabolic phenotype can be achieved but structural information can be obtained by MS and MS/MS, which can further be used to identify potential biomarkers[7]. Therefore, MS–based strategies are characterized by its high sensitivity, improved metabolite discrimination, coverage of the metabolome, and modularity to perform compound–class–specific analysis.

Nevertheless, among the main drawbacks of this technique (and in general all MS–hyphenated techniques) is the generation of very complex datasets. This is the result of: (*i*) the noisy MS background from biological sample matrices and impurities in the mobile phases; (*ii*) the existence of multiple features from a single metabolite due to protonated, adduct, neutral loss and fragment ions in case of ESI

ionization is used; (*iii*) poor separation of structural isomers; (*iv*) co–elution of multiple components at a given retention time; (*vi*) the inevitable RT and *m/z* shifts over time; and (*vii*) the wide dynamic ranges of individual metabolites across all the samples[29].

In addition, critical issues when obtaining data from LC–MS data are robustness and reproducibility; special care and awareness should always be taken when an experiment is being carried out. LC parameters such as curve pressure, which should be stable and not significantly changing among runs, and chromatograms inspections, at the shape of the profile and peak shift, signal intensity, mass accuracy calibration of the reference masses, and if possible, standard mix should always be taken under consideration.

When an experiment of non–target analysis is performed, the instrumental stability must be guaranteed. The way this is done is through the use of quality control (QC) samples. Usually a QC sample is prepared by mixing small aliquots (e.g., 10–20µl) of each sample in the study set, thus providing a true representation of most metabolites present in the sample set. The QC sample should be injected before the study samples, during the sequence (normally each 5 runs) and at the end of the work list. Of note, the first few injections of any type of biological sample often give unrepresentative results, mainly due to small changes in both chromatographic RT and signal intensity. Therefore, running five to ten QC samples at the start of the sequence is compulsory[7]. Thus, both QC samples and standard mix enable the researcher to obtain information about the system suitability and stability, highlighting any problems coming from system contamination or instrument failure. It is often observed that changes in analytical performance over the analysis time occur from the contamination of the MS ion source, which leads to changes in instrument sensitivity. This can be sample–type dependent, and therefore randomization of samples is important. Table 5.5 summarizes the strengths and drawbacks of LC–MS.

Table 5. Relative strengths and weakness of LC–MS for metabolomics studies[7, 30].

Strengths	Drawbacks
▪ Highly sensitivity	▪ Generation of very complex datasets
▪ Very selective (especially in MS^n)	▪ Quantification requires internal standard
▪ Suitable for quantitative target analysis and non–target analysis	▪ Destructive, particularly complicated for analysis of valuable samples
▪ Structural information throughout isotopic distribution and fragmentation patterns (MS^n)	▪ Limited to liquid samples or extracts isolated from solid samples (such as tissue)
▪ Accurate *m/z* measurements ideal for identification of metabolites	▪ Laborious sample preparation protocols
▪ Excellent resolution (improved by the LC separation)	▪ Need to use QC samples and standard mix for checking system suitability and stability
▪ Samples can be measured in positive and negative mode	▪ Detection depends on ionization efficiency (ESI present high potential for ion suppression)
▪ Analysis and detection of a wide range of metabolite classes	▪ Low analytical performance and stability over large dataset analysis
▪	▪ Expensive

3.5. Insight into the application of LC–QTOF–MS

The methods based on ^{1}H–NMR spectroscopy and MS remain the dominant analytical platforms for non–target analysis[7]. Indeed, in the case of LC–MS, it was not easy to establish a proper work methodology. The researcher that is working with LC–MS instrumentation requires expertise in the field. It is very important to continuously check the key characteristics of the analytical system (all mentioned before), and it is always better to be sure about the instrument stability than to later discover that the experiment was a waste of money and samples.

In the case of this article, the patients involved in this study were treatment–resistant bipolar depressive (BD) and they were using two different mood stabilizers (VPA and lithium) as co–medication, apart from placebo or ketamine infusion. This suggested to us the possibility that analysis could obtain

different results (peak shift, chromatogram shape, ionization, etc.) so special care in sample randomization and instrumental conditions was taken.

One of the main aspects of LC–MS is the data treatment. After the data from an experiment is obtained, the data then is reprocessed. Several software programs were developed for obtaining the chemical entities from raw data (Chapter 1). LC–MS generates complex data generally rich in noise and useless and random information. So, the initial step is to always filter the raw data. In any MS–based experiment, we expect to work with reliable information to be statistically compared afterwords. One of the ways to assert this objective is keeping only those features that have a coefficient of variation (CV %) lower than 20-35% from the QCs samples. This is one of the main approaches for data filtration[31]. The data matrix after filtration is examined through a PCA plot of all datasets (biological and QC samples) and confirmed by the observation of QCs samples clustering. Then the dataset can be used further for multivariate statistical analysis. It is very important to be very clear with the objective pursued and the groups that are going to be compared. For this project, our gold objective was the possibility to find biomarkers that could predict the response (or not) to the medical treatment (ketamine infusion) regardless of the co–medication (VPA and lithium). Our findings suggested that the significant masses for response to ketamine inside each group, and combining the samples from both, where very different (Figure 2C of the publication[32]). This made us to work more in depth with the lithium samples, so for this again only Li samples and QCs samples were aligned, masses were filtered and suitability of data was tested on a PCA plot. This is just an example of the complexity of the data treatment from LC–MS.

Another aspect of LC–MS data is the amount of data that is handled. For example, for the lithium samples from ketamine infusion arm, the initial amount of features was 18,926 and after filtering based on QCs samples keeping the features with a CV less than 35%, and present in at least 100% of one of the two groups. In that case the features were reduced to 885 (less than a half number from the initial). Later, after multivariate analysis, 165 masses were selected and from these only 18 were identified by concordant RT for the LC method employed: mass accuracy, isotope distribution and MS/MS fragmentation pattern. Finally from these 18, only one standard reagent was available and the metabolite was totally confirmed (Figure 5.6).

In brief, the abundance of data is not in itself a guarantee of obtaining useful information in an investigated system. On the contrary, data from the metabolomics field need to be processed, and correctly analysed in order to highlight useful information among the measurements[33]. For this research work the total amount of comparisons was 6, making it difficult to identify all the discriminant metabolites, forcing focus onto only one of high interest.

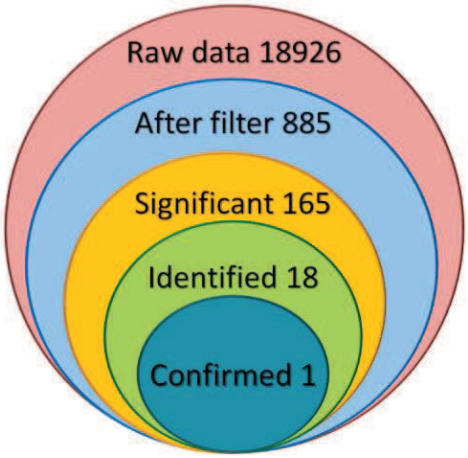


Figure 5.6. Graphical presentation of the impact of each data analysis step used for LC–MS data in the comparison of Rs and NRs taking lithium in the ketamine arm of the therapy

When working with such high amount of features or variables, it is necessary to confirm the significance of the biomarkers by other statistical parameters that contribute to picture the importance of this metabolite for the comparison, i.e. univariate analysis (usually by *t*–test or ANOVA) and percentage of change (%) of the metabolite.

Also is worth mentioning additional challenges in the data treatment such as missing values, logarithmic transformation, and normalization algorithms. MS–based data is typically characterized by missing values in the data set (even after data filtration) which can be the result of a fail in the data reprocess algorithm (most common cause), low concentration near the limit of detection or zones with high noise, ion suppression, or simply because that particularly sample has genuine biological reasons that it is below the limit of the detection. There is no actual protocol or procedure that talks about this issue and how to handle it, although it has been proposed various algorithms for the estimation of missing data[34].

It has been proved that the logarithmic transformation of MS data is needed to correct the drift in the signal of LC–MS[35–36]. Logarithmic transformation makes the data distribution more symmetric, and for multivariate models this has a pseudo-scaling effect because it reduces large values in the data set relatively more than the small values[36]. This helps in fitting the model and finding the metabolites with high covariance and correlation with the class. For univariate analysis, e.g., in the case of *t*–test, any calculation made using logarithmic data of a significant metabolite will suppose that the statistical difference is not in the arithmetic average but in the geometric mean, and this should be studied in more depth. In my particular experience, the *p* value of highly significant metabolites will stand from raw data or after logarithmic transformation.

Data normalization is highly important for biofluids with significant concentration variations, such as urine which varies as a result of fluids intake. There is not a standard algorithm for MS–base data, although Godzien, J., 2013[37]suggests that 24h urine volume could be a good normalization parameter. Also Veselkov, K, 2011[35]showed that median fold change normalization was the best algorithm for correct urine dilution.

REFERENCES.

1. Theodoridis, G. A.; Gika, H. G.; Want, E. J.; Wilson, I. D., Liquid chromatography–mass spectrometry based global metabolite profiling: a review. *Anal Chim Acta* **2012,***711*, 7–16.
2. Snyder, L. R.; Kirkland, J. J.; Dolan, J. W., *Introduction to Modern Liquid Chromatography*. 3 ed.; John Wiley & Sons, Inc.: United States of America, **2010**; p 960
3. Niessen, W. M. A., *Liquid Chromatography–Mass Spectrometry*. 3 ed.; CRC Press. Taylor & Francis Group.: United States of America, **2006**.
4. Rojo, D.; Barbas, C.; Rupérez, F. J., LC–MS metabolomics of polar compounds. *Bioanalysis* **2012,***4* (10), 1235–43.

5. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt–Kopplin, P., Liquid chromatography–mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **2013,***1292*, 51–65.

6. Sandra, K.; Pereira Ados, S.; Vanhoenacker, G.; David, F.; Sandra, P., Comprehensive blood plasma lipidomics by liquid chromatography/quadrupole time–of–flight mass spectrometry. *J Chromatogr A* **2010,***1217* (25), 4087–99.

7. Want, E. J.; Masson, P.; Michopoulos, F.; Wilson, I. D.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Loftus, N.; Holmes, E.; Nicholson, J. K., Global metabolic profiling of animal and human tissues via UPLC–MS. *Nature protocols* **2013,***8* (1), 17–32.

8. Spagou, K.; Wilson, I. D.; Masson, P.; Theodoridis, G.; Raikos, N.; Coen, M.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Nicholson, J. K.; Want, E. J., HILIC–UPLC–MS for exploratory urinary metabolic profiling in toxicological studies. *Anal Chem* **2011,***83* (1), 382–90.

9. Callahan, D. L.; De Souza, D.; Bacic, A.; Roessner, U., Profiling of polar metabolites in biological extracts using diamond hydride–based aqueous normal phase chromatography. *Journal of separation science* **2009,***32* (13), 2273–80.

10. Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G., Evaluation of coupling reversed phase, aqueous normal phase, and hydrophilic interaction liquid chromatography with Orbitrap mass spectrometry for metabolomic studies of human urine. *Anal Chem* **2012,***84* (4), 1994–2001.

11. Ciborowski, M.; Javier Ruperez, F.; Martinez–Alcazar, M. P.; Angulo, S.; Radziwon, P.; Olszanski, R.; Kloczko, J.; Barbas, C., Metabolomic approach with LC–MS reveals significant effect of pressure on diver's plasma. *J Proteome Res* **2010,***9* (8), 4131–7.

12. Holcapek, M.; Jirasko, R.; Lisa, M., Recent developments in liquid chromatography–mass spectrometry and related techniques. *J Chromatogr A* **2012,***1259*, 3–15.

13. Kind, T.; Fiehn, O., Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical reviews* **2010,***2* (1–4), 23–60.

14. Khan, M. T. H.; Ather, A., Metabolomics – systematic studies of the metabolic profiling. In *Lead Molecules from Natural Products: discovery and new trends.*, Elsevier: The Netherlands., 2006; p 9.

15. Katajamaa, M.; Oresic, M., Processing methods for differential analysis of LC/MS profile data. *BMC bioinformatics* **2005,***6*, 179.

16. Lu, X.; Zhao, X.; Bai, C.; Zhao, C.; Lu, G.; Xu, G., LC–MS–based metabonomics analysis. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2008,***866* (1–2), 64–76.

17. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry–based metabolomics. *Mass spectrometry reviews* **2007,***26* (1), 51–78.

18. Becker, S.; Kortz, L.; Helmschrodt, C.; Thiery, J.; Ceglarek, U., LC–MS–based metabolomics in the clinical laboratory. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2012,***883–884*, 68–75.

19. Laganowsky, A.; Reading, E.; Hopper, J. T.; Robinson, C. V., Mass spectrometry of intact membrane protein complexes. *Nature protocols* **2013,***8* (4), 639–51.

20. Theodoridis, G.; Gika, H. G.; Wilson, I. D., LC–MS–based methodology for global metabolite profiling in metabonomics/metabolomics. *Trends in Analytical Chemistry* **2008,***27* (3), 10.

21. (a) Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC–MS. *Nature protocols* **2010,***5* (6), 1005–18; (b) Chambers, E.; Wagrowski–Diehl, D. M.; Lu, Z.; Mazzeo, J. R., Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2007,***852* (1–2), 22–34.

22. Ens, W.; Standing, K. G., Hybrid quadrupole/time–of–flight mass spectrometers for analysis of biomolecules. *Methods in enzymology* **2005,***402*, 49–78.

23. Glish, G. L.; Burinsky, D. J., Hybrid mass spectrometers for tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **2008,***19* (2), 161–72.

24. Wells, J. M.; McLuckey, S. A., Collision–induced dissociation (CID) of peptides and proteins. *Methods in enzymology* **2005,***402*, 148–85.

25. Milne, S.; Ivanova, P.; Forrester, J.; Alex Brown, H., Lipidomics: an analysis of cellular lipids by ESI–MS. *Methods* **2006,***39* (2), 92–103.

26. Meikle, P.; Barlow, C.; Weir, J., Lipidomics and Lipid Biomarker Discovery. *Australian Biochemist.* **2009,***40* (3), 5.

27. Cui, Z.; Thomas, M. J., Phospholipid profiling by tandem mass spectrometry. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **2009,***877* (26), 2709–15.

28. German, J. B.; Hammock, B. D.; Watkins, S. M., Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **2005,***1* (1), 3–9.

29. Han, J.; Datla, R.; Chan, S.; Borchers, C. H., Mass spectrometry–based technologies for high–throughput metabolomics. *Bioanalysis* **2009,***1* (9), 1665–84.

30. Lao, Y. M.; Jiang, J. G.; Yan, L., Application of metabonomic analytical techniques in the modernization and toxicology research of traditional Chinese medicine. *British journal of pharmacology* **2009,***157* (7), 1128–41.

31. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis–McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; Human Serum Metabolome, C., Procedures for large–scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature protocols* **2011,***6* (7), 1060–83.

32. Villasenor, A.; Ramamoorthy, A.; Silva Dos Santos, M.; Lorenzo, M. P.; Laje, G.; Zarate, C., Jr.; Barbas, C.; Wainer, I. W., A pilot study of plasma metabolomic patterns from patients treated with ketamine for bipolar depression: evidence for a response-related difference in mitochondrial networks. *Br J Pharmacol* **2014,** *171* (8), 2230-42.

33. Eriksson, L.; Antti, H.; Gottfries, J.; Holmes, E.; Johansson, E.; Lindgren, F.; Long, I.; Lundstedt, T.; Trygg, J.; Wold, S., Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal Bioanal Chem* **2004,***380* (3), 419–29.

34. Hrydziuszko, O.; Viant, M. R., Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics.* **2011,***8*, 14.

35. Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier–Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., Optimized preprocessing of ultra–performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem* **2011,***83* (15), 5864–72.

36. van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006,***7*, 142.

37. Godzien, J.; Ciborowski, M.; Angulo, S.; Barbas, C., From numbers to a biological sense: How the strategy chosen for metabolomics data treatment may affect final results. A practical example based on urine fingerprints obtained by LC–MS. *Electrophoresis* **2013,***34* (19), 2812–26.

British Journal of Pharmacology

DOI:10.1111/bph.12494
www.brjpharmacol.org

Themed Issue: Mitochondrial Pharmacology: Energy, Injury & Beyond

RESEARCH PAPER

A pilot study of plasma metabolomic patterns from patients treated with ketamine for bipolar depression: evidence for a response-related difference in mitochondrial networks

A Villaseñor[1], A Ramamoorthy[2], M Silva dos Santos[1,3], M P Lorenzo[1], G Laje[4,5,6], C Zarate Jr[4], C Barbas[1] and I W Wainer[2]

[1]*Center for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Madrid, Spain,* [2]*Laboratory of Clinical Investigation, Intramural Research Program, National Institute on Aging, National Institutes of Health (NIH), Baltimore, MD, USA,* [3]*Faculty of Pharmacy, Federal University of Minas Gerais, Belo Horizonte, Brazil,* [4]*Experimental Therapeutics & Pathophysiology Branch Intramural Research Program, National Institute of Mental Health (NIMH), NIH, Bethesda, MD, USA,* [5]*Human Genetics Branch and Genetic Basis of Mood and Anxiety Disorders Section, Intramural Research Program, NIMH, NIH, Bethesda, MD, USA, and* [6]*Maryland Institute for Neuroscience and Development (MIND), Bethesda, MD, USA*

Correspondence
Irving W Wainer, Laboratory of Clinical Investigation, National Institute on Aging, NIH, 251 Bayview Boulevard, 08B133, Baltimore, MD 21224, USA.
E-mail: wainerir@grc.nia.nih.gov

Keywords
mitochondrial function; fatty acid metabolism; pharmacometabolomics; lithium; valproate

Received
8 August 2013
Revised
14 October 2013
Accepted
28 October 2013

BACKGROUND AND PURPOSE

(R,S)-ketamine produces rapid and significant antidepressant effects in approximately 65% of patients suffering from treatment-resistant bipolar depression (BD). The genetic, pharmacological and biochemical differences between ketamine responders and non-responders have not been identified. The purpose of this study was to employ a metabolomics approach, a global, non-targeted determination of endogenous metabolic patterns, to identify potential markers of ketamine response and non-response.

EXPERIMENTAL APPROACH

Plasma samples from 22 BD patients were analyzed to produce metabolomic patterns. The patients had received ketamine in a placebo-controlled crossover study and the samples were obtained 230 min post-administration at which time the patients were categorized as responders or non-responders. Matching plasma samples from the placebo arm of the study were also analysed. During the study, the patients were maintained on either lithium or valproate.

KEY RESULTS

The metabolomic patterns were significantly different between the patients maintained on lithium and those maintained on valproate, irrespective of response to ketamine. In the patients maintained on lithium, 18 biomarkers were identified. In responders, lysophosphatidylethanolamines (4) and lysophosphatidylcholines (9) were increased relative to non-responders.

CONCLUSIONS AND IMPLICATIONS

The results indicate that the differences between patients who respond to ketamine and those who do not are due to alterations in the mitochondrial β-oxidation of fatty acids. These differences were not produced by ketamine administration. The data indicate that pretreatment metabolomics screening may be a guide to the prediction of response and a potential approach to the individualization of ketamine therapy.

LINKED ARTICLES
This article is part of a themed issue on Mitochondrial Pharmacology: Energy, Injury & Beyond. To view the other articles in this issue visit http://dx.doi.org/10.1111/bph.2014.171.issue-8

Abbreviations
5-HT, 5-hydroxytryptamine; BD, bipolar depression; LC-QTOF-MS, liquid chromatography coupled to QTOF-MS; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; MADRS, Montgomery Åsberg Depression Rating Scale; MDD, major depressive disorder; MFE, molecular feature extraction algorithm; MG, monoglycerides; NR, non-responder; OPLS-DA, orthogonal partial least square discriminant analysis; PCA, principal component analysis; PLA, phenyl lactic acid; PLS-DA, partial least square discriminant analysis; PTSD, post-traumatic stress disorder; Q^2, prediction power score; QC, quality control; QTOF, quadrupole time-of-flight mass spectrometer; R, responder; R^2Y, classification score; SAM, S-adenosyl-methionine; SNRI, serotonin-noradrenaline reuptake inhibitors; SSRI, selective 5-HT reuptake inhibitors; TML, N_6,N_6,N_6-trimethyl-L-lysine.

Introduction

The majority of the mood stabilizers and antidepressants used to treat bipolar depression (BD) target the 5-hydroxytryptaminergic and noradrenergic systems (Machado-Vieira *et al.*, 2008). These drugs are associated with a considerable lag in the onset of antidepressant action, high inter-individual variability in response and many 'treatment-resistant' patients (Machado-Vieira *et al.*, 2008). However, recent data have indicated that there is an alternative approach to the treatment of BD based upon the targeting of the glutamatergic system using antagonists of the NMDA receptor (Zarate *et al.*, 2006; 2012). A number of these studies have utilized a subanesthetic dose of the NMDA receptor antagonist (R,S)-ketamine and observed a rapid (within 4 h), and significant antidepressant effect in approximately 65% of previously treatment-resistant BD patients (Zarate *et al.*, 2006, 2012; Dolgin, 2013). The genetic, pharmacological and biochemical differences between ketamine-responders (Rs) and non-responders (NRs) have not been identified. The objective of this pilot study is to employ a metabolomics approach to identify potential markers of ketamine response and non-response.

Metabolomics, a global, non-targeted approach to the study of biochemical processes and metabolic networks (Kaddurah-Daouk *et al.*, 2008), has been used to identify disease-specific metabolic profiles and biomarkers of CNS disorders, including major depressive disorder (MDD) (Schwarz and Bahn, 2008; Kaddurah-Daouk and Krishnan, 2009; Quinones and Kaddurah-Daouk, 2009). Recent studies in MDD patients have also investigated changes in metabolomic patterns produced by administration of selective 5-hydroxytryptamine (5-HT) reuptake inhibitors (SSRI) and pharmacogenetic analysis to identify glycine and glycine dehydrogenase as citalopram/escitalopram response makers (Ji *et al.*, 2011; Abo *et al.*, 2012). Metabolomic studies have demonstrated that fatty acid metabolism is significantly lower and shifted from β-oxidation to ω-oxidation in depressed patients, compared with non-depressed controls (Maes *et al.*, 1996; Paige *et al.*, 2007; Steffens *et al.*, 2010), indicating that mitochondrial function may be associated with the disease state. These results are consistent with the association of mitochondrial function with depression and anxiety (Burroughs and French, 2007) and with MDD and BD-related changes in phospholipid metabolism (Modica-Napolitano and Renshaw, 2004). In addition, stress-induced changes in mitochondria membrane potential have been suggested as a mechanism for hippocampus atrophy in posttraumatic stress disorder (PTSD) (Zhang *et al.*, 2006). *Postmortem* studies in patients with PTSD identified 119 dysregulated genes, a number of these genes are associated with mitochondrial dysfunction and oxidative phosphorylation (Su *et al.*, 2008).

In the current pilot study, a metabolomic analysis was performed using plasma samples obtained from 22 treatment-resistant patients with BD who had received ketamine in a placebo-controlled crossover study (Diazgranados *et al.*, 2010; Zarate *et al.*, 2012). The samples were obtained 230 min post-ketamine administration, at which time the patients were categorized as Rs or NRs. Metabolomic global profiling was also performed in 17 patients using matching plasma samples from the placebo arm of the study. During the study, the patients were maintained on either lithium or valproate. The effects of these mood stabilizers on ketamine response patterns were also examined as previous studies of *post mortem* brain tissues from BD patients and rat studies have indicated that lithium and valproate have different effects on the levels of excitatory/inhibitory neurotransmitters (Lan *et al.*, 2009).

The results demonstrate that the metabolomic patterns observed in the patients receiving lithium (lithium-subgroup) were significantly different from those obtained with the valproate-subgroup, irrespective of the response to ketamine. Within the lithium-subgroup, 18 compounds were found to be significantly different between patients who responded to ketamine (Rs) and non-responders (NRs), with 16 of the 18 associated with the metabolism of fatty acids. Relative increases or decreases were observed in lysophosphatidylcholines (9 compounds), lysophosphatidylethanolamines (4), monoglycerides (2) and the carnitine precursor N_6,N_6,N_6-trimethyl-L-lysine. Thus, the data suggest that key factors in the clinical response or non-response to ketamine in BD patients are differences in mitochondrial function reflected in fatty acid metabolism.

Methods

Patient selection and ketamine administration

The Combined Neuroscience Institutional Review Board of the National Institutes of Health approved the study. All subjects provided written informed consent and were studied as inpatients at the National Institute of Mental Health Clinical Research Center, Mood Disorders Research Unit in Bethesda, Maryland. Plasma samples were obtained from 22 BD patients enrolled in a placebo-controlled study of the effect of ketamine on depression (Diazgranados *et al.*, 2010). At the time of the study, the patients were experiencing a major depressive episode without psychotic features and had been maintained on a mood stabilizer either lithium or valproate for 4 weeks before the trial and during the ketamine infusion. Lithium and valproate were administered twice a day (morning and evening) and the blood levels of the agents were within therapeutic range, serum lithium, 0.6–1.2 $mEq{\cdot}L^{-1}$, serum valproate 50–125 $\mu g{\cdot}mL^{-1}$, throughout the study. No other psychotropic medications were permitted in the 2 weeks before and during the study.

The patients received a single i.v. infusion of 0.5 $mg{\cdot}kg^{-1}$ of ketamine hydrochloride over the course of 40 min, and symptoms, including Montgomery Åsberg Depression Rating Scale (MADRS) scores and blood samples, were collected into heparinized tubes at 40 min (the end of the infusion), 80, 110 and 230 min post-infusion, and at day 1. The differential in MADRS scores from 0 to 230 min were used to delineate Rs and NRs with a relative improvement of ≥50% in MADRS score signifying a response to ketamine.

Metabolomics assay of plasma samples using LC-QTOF-MS

The experimental procedures utilized in the metabolomics study involved the sequential statistical treatment of data obtained from the analysis of the experimental and control samples. The samples were analyzed using liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer (LC-QTOF-MS), which initiated the detection, quantitation and identification of the unique biochemicals associated with the observed clinical response (Figure 1).

In the initial step, a frozen plasma sample was thawed on ice, 3 volumes of ice-cold methanol : ethanol (1:1, v/v) was added to 1 volume of plasma, the mixture vortexed for 1 min, placed on ice for 5 min and then centrifuged at 15 700 g for 20 min at 4°C. The supernatant was collected and filtered through a 0.22 μm nylon filter. Quality Control (QC) samples were prepared by pooling an aliquot of each of the filtered supernatant. The samples were analysed by liquid chromatography following a previously described approach (Ciborowski *et al.*, 2012). All chromatographic separations were performed using a 1200 series HPLC (Agilent Technologies, Waldbronn, Germany), Supelco Discovery HS C18 analytical column (15 cm × 2.1 mm, 3 μm) and guard column (2 cm × 2.1 mm, 3 μm; Sigma Aldrich, Steinheim, Germany), the autosampler and columns were maintained at 4 and 40°C, respectively, and the injection volume was set to 10 μL. Data were collected in positive electrospray ionization mode in separate runs on a QTOF (Agilent 6520) operated in full scan mode from 100 to 1000 m/z. During the analysis, two reference masses: 121.0509 m/z ($C_5H_4N_4$) and 922.0098 m/z ($C_{18}H_{18}O_6N_3P_3F_{24}$), were continuously measured to allow constant mass correction. The capillary voltage was 3000 V with a scan rate of 1.02 scan per second and the nebulizer gas flow rate was 10.5 $L{\cdot}min^{-1}$. LC-MS grade acetonitrile, methanol, ethanol, formic acid and standards used were purchased from Sigma-Aldrich. Ultra-pure water Milli-Q Water System (Millipore, Billerica, MA, USA) was used for mobile phase and all standard solutions.

LC-QTOF-MS data processing and analysis

Data were re-processed using molecular feature extraction (MFE) tool using the Mass Hunter Qualitative Analysis software B.04.00 (Agilent Technologies) which allowed subtraction of background noise and data reduction. The MFE file gave a list of each mass and retention time pairs with associated intensities for all detected peaks. Alignment and filtering of the primary data were performed on Mass Profiler Professional B.02.00 (Agilent Technologies) software. Masses in the samples that were not present in 100% of participants in at least one group and that had a coefficient of variation above 35% in QC samples were filtered out.

Biomarker discovery

The masses were exported to SIMCA-P+ 12.0 (Umetrics, Umeå, Sweden) for multivariate statistical analysis. Principal component analysis (PCA), partial least square discriminant analysis (PLS-DA) and orthogonal PLS-DA (OPLS-DA) models were used to discriminate the samples. Potential biomarkers were selected through an OPLS-DA model. Differences in metabolites were assessed through the loading column plot where an error bar (jack-knife) was calculated for each variable. Variables with a 95% jack-knife confidence level were selected for further identification.

An external cross-validation test was used to verify the predictability of the PLS-DA model and avoid the risk of overfitting (Rubingh *et al.*, 2006). The sample set was randomized and split in three groups. In each prediction set, one group was excluded and predicted by the remaining groups. The procedure was repeated until the three groups were predicted and the global percentage of samples classified correctly was calculated.

Compound identification

The databases METLIN, LIPID MAPS, MASSTRIX and HMDB were searched for hits against the identified discriminant accurate masses. For each hit, the proposed formula was compared with the experimental isotopic pattern distribution. To confirm the identity of statistically significant compounds, LC-MS/MS analyses were performed on a QTOF (model 6520, Agilent Technologies) using the initial chromatographic conditions. Ions were targeted based on previously determined mass and retention time, nitrogen was used as the collision gas, and collision energy was adjustable with slope of 3.6 V/100 Da and offset 4.8 V for fragmentation. Compound identification was performed as previously described (Lin *et al.*, 2010) – in brief, the corresponding molecular ion (m/z) was

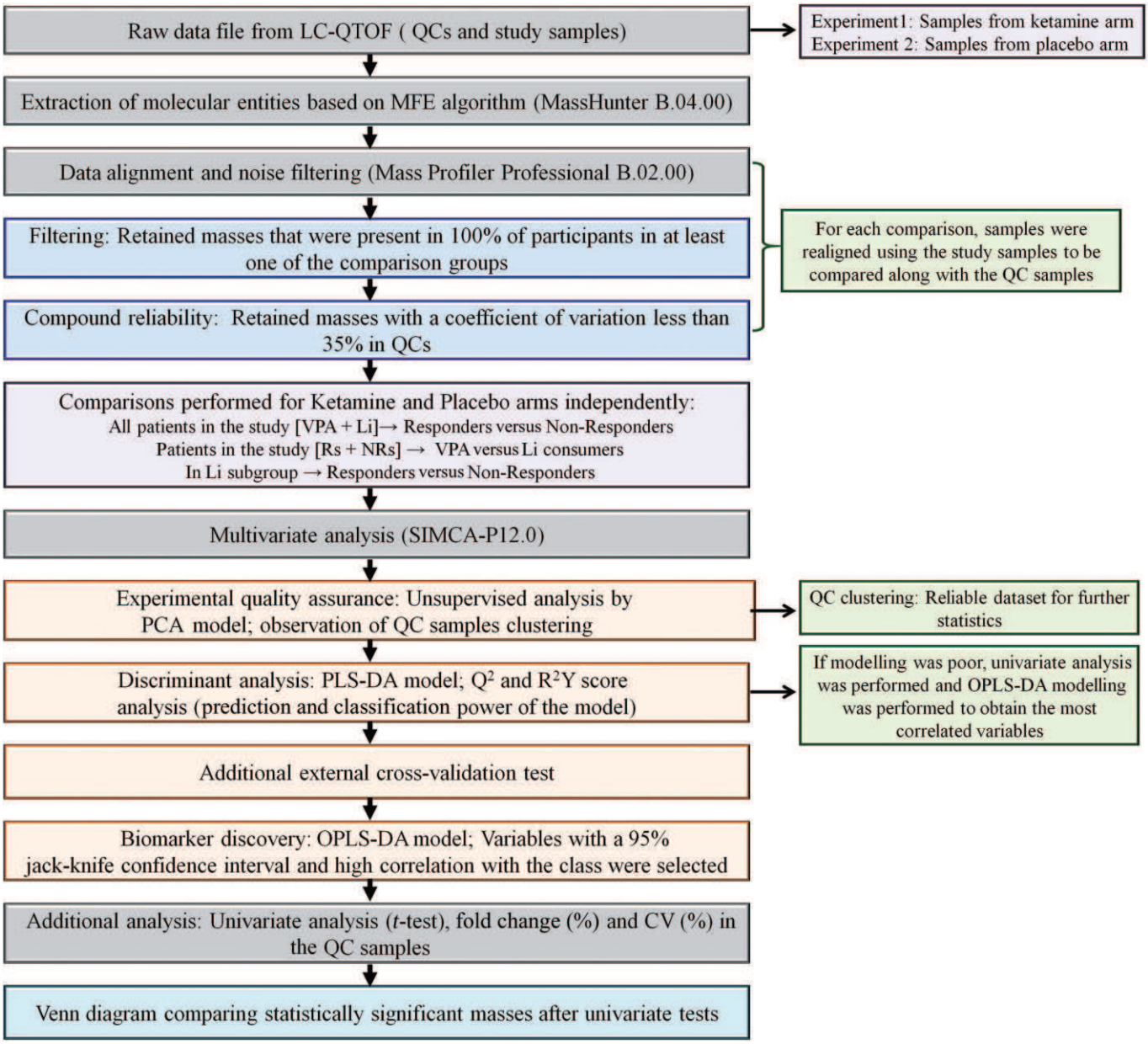


Figure 1

Analysis of metabolomics data from plasma samples. A schematic representation of the sample analysis including detection, quantitation and identification of the unique biochemicals associated with response/non-response to ketamine therapy in bipolar depression.

searched as extracted ion chromatogram and according to its retention time. Elemental composition of the peak (chemical formula) based on the exact mass and isotope pattern recognition was compared with the database hit considering the probability score. If available, the MS/MS spectra were compared with spectra in the MS/MS spectra library (METLIN). For compounds whose fragmentation pattern was not present in METLIN database, the patterns were predicted using ACD/ChemSketch software v.12.01 (ACD/Labs, Toronto, ON, Canada). When possible, the identity of the compound was confirmed using a commercially available reagent.

Results

Demographics and treatment characteristics of the patient population

The patient samples (n = 22) analysed in the study were obtained at 230 min post-initiation of ketamine or placebo infusion (Diazgranados *et al.*, 2010). The MADRS score at 230 min was compared with the pre-infusion scores and patients with a ≥50% reduction were classified as Rs, and, using the criteria, this study included 13 Rs (67 ± 12% reduc-

Table 1

Clinical characteristics and demographics of the BD patients included in this study

	Combined		Patients taking lithium		Patients taking valproate	
Characteristics	**Responders**	**Non-responders**	**Responders**	**Non-responders**	**Responders**	**Non-responders**
Sample size (*N*)	13	9	9	7	4	2
Age (years) mean ± SD	42.3 ± 10.6	50.9 ± 10.2	40.6 ± 11.1	50.6 ± 10.2	46.3 ± 9.5	52 ± 14.1
Gender (% female)	69.2	66.7	77.8	57.1	50	100
Race (% Caucasian)	100	88.9	100	85.7	100	100
MADRS score at 230 min mean ± SD	11 ± 4	29 ± 7	12 ± 4	28 ± 7	8 ± 5	32 ± 1
Percentage reduction in MADRS mean ± SD	67 ± 12	13 ± 13	64 ± 11	14 ± 15	75 ± 13	9

BD, bipolar depression; MADRS, Montgomery Åsberg Depression Rating Scale.

tion) and 9 NRs (13 ± 13% reduction). Sixteen of the patients were maintained on lithium and 6 on valproate (Table 1).

Global pharmacometabolomic profiles

Chromatograms from the analysis of the samples obtained from 22 patients post- ketamine infusion and 6 QC samples were multi-aligned and data were filtered to remove noise. The quality of the global profiling analysis was evaluated using PCA method. The results indicated that the QCs clustered tightly (Figure 2A), suggesting that the dataset can be used for further statistical analysis (Milne *et al.*, 2006). Samples were categorized as coming from R or NRs as described above. The dataset was analysed by PLS-DA and OPLS-DA. Modelling the whole dataset presented both reasonable and poor predictive power ($Q^2 > 0.4$ and 0.2 respectively; data not shown). The *t*-test analysis indicated that 75 metabolites were statistically significant ($P < 0.05$). Using these metabolites, an OPLS-DA model was generated. The OPLS-DA was jack-knifed (Figure 2B) and the result suggests that the 52 masses that were selected predicted the metabolic variability observed between the Rs and NRs groups. Putative metabolite identification indicated that 11 metabolites were up-/down-regulated in NRs, with the majority of the compounds identified as lipids and fatty acid acyl derivatives (Table 2). Global profiling was also performed on 17 plasma samples obtained from the placebo arm of the study. The placebo samples were classified as obtained from Rs ($n = 10$) or NRs ($n = 7$) based on the observed response in the ketamine arm of the study. Multivariate analysis of the placebo samples did not show any model fitting. However, univariate analysis revealed that in the patients categorized as NRs, octanoyl- and decanoylcarnitines were increased while carnitine and two glycerophospholipids (16:1 and 19:1) were decreased (Supporting Information Table S1).

Influence of lithium and valproate on global pharmacometabolomic profiles

In order to determine the influence of mood stabilizer on the response to ketamine therapy, the Rs and NRs within lithium and valproate subgroups were compared independently by univariate *t*-test analysis. Significant masses from this analysis were compared with masses identified from *t*-test analysis of the combined (lithium + valproate) group. The results indicate that in the ketamine arm of the study, 28 metabolites were significantly different between patients taking lithium and those receiving valproate (Supporting Information Table S2). Statistically significant masses ($P < 0.05$) from each of the comparison are displayed in a Venn diagram (Figure 2C). Only two masses were common between the lithium- and the valproate-subgroups, and only one mass was common among the three comparisons, suggesting that the mood stabilizers influenced the metabolomic patterns. These differences were also observed in the samples obtained from the placebo arm of the study (Supporting Information Table S3), indicating that the differences in the metabolomic patterns were associated with the administration of valproate and lithium rather than ketamine. In both the ketamine and placebo arms of the study, the concentrations of acylcarnitines, octenoic acid and its analogue metabolites were increased in the valproate-subgroup relative to the lithium-subgroup. In addition, there were relative increases in phenylalanine and tryptophan in the lithium-subgroup, which appears to be due to the action of ketamine, independent of patient response. R^2Y and Q^2 parameters for these models showed high-quality sample classification and good predictive power (Supporting Information Figure S1).

Metabolomic profile of patients taking lithium as the mood stabilizer

The initial objective of the study was to determine the global metabolomic markers associated with the clinical

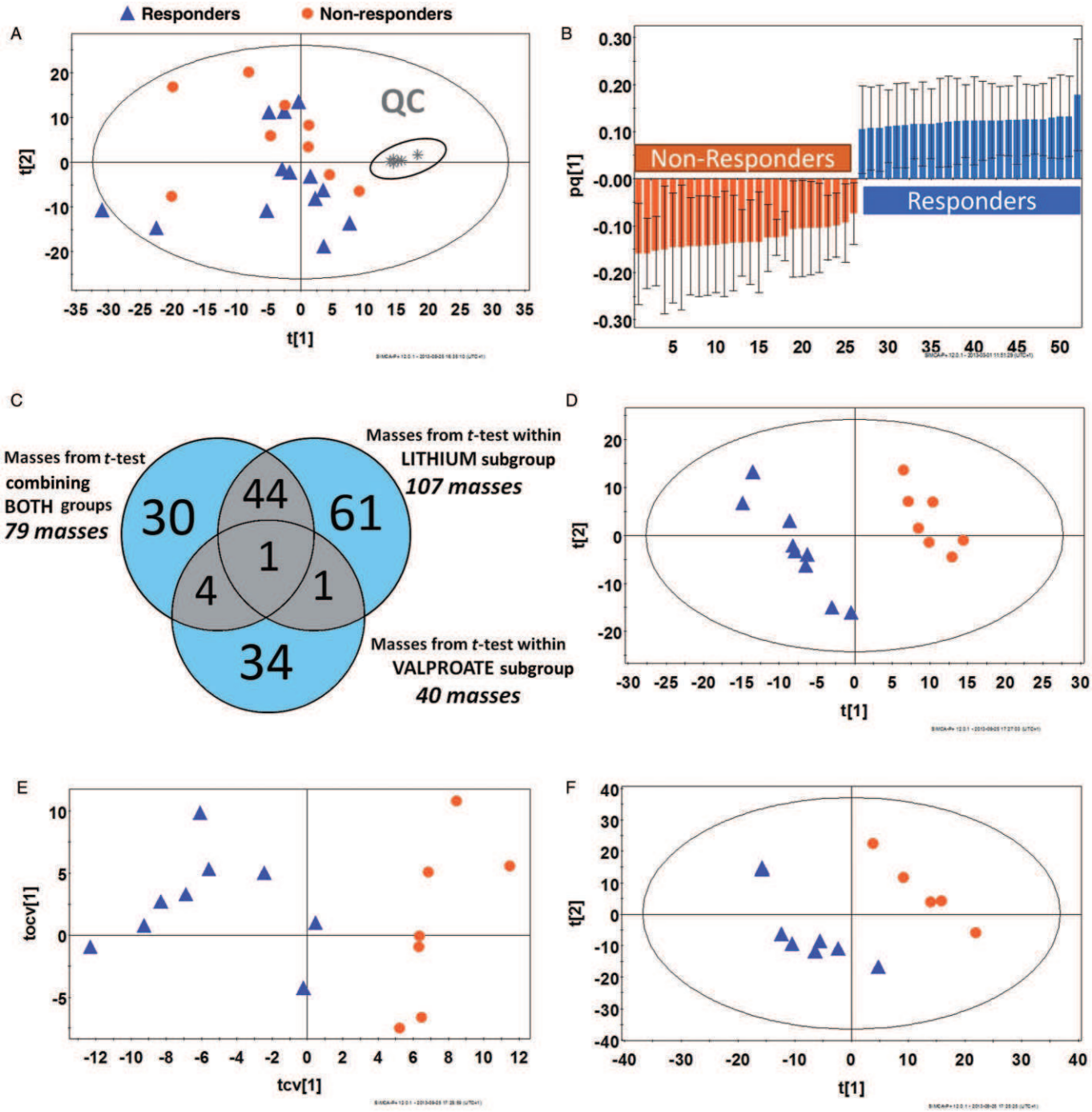


Figure 2

Statistical analysis of global profiles derived from bipolar patients treated with ketamine or placebo. (A) PCA model of LC-MS dataset after filtration (970 out of 18926 features) from 22 patients with BD after 230 min post-infusion of ketamine showing the clustering of QC samples (shown in the gray oval) [$R^2Y = 0.34$; $Q^2 = 0.09$]. (B) Covariance plot from OPLS-DA model for metabolites different between Rs and NRs after *t*-test including jack-knife (confidence interval >95%) [$R^2Y = 0.98$; $Q^2 = 0.85$; $n = 75$ features]. (C) Venn diagram displaying overlapping metabolites after *t*-test for the comparison of Rs versus NRs to ketamine therapy in lithium and valproate subgroups individually and when combined together. (D) PLS-DA plot of plasma samples in patients taking lithium in the ketamine arm of the therapy, Rs ($n = 9$) against NRs ($n = 7$) using the filtered dataset [$R^2Y = 0.997$; $Q^2 = 0.672$; $n = 885$ features]. (E) OPLS-DA cross-validated plot of plasma samples in patients taking lithium in the ketamine arm of the therapy, Rs ($n = 9$) against NRs ($n = 7$) using the filtered dataset [$R^2Y = 1.000$; $Q^2 = 0.779$; $n = 885$ features]. (F) PLS-DA model of Rs ($n = 8$) against NRs ($n = 5$) from patients taking lithium in the placebo arm [$R^2Y = 0.949$; $Q^2 = 0.0.539$; $n = 1325$ features]. UV scaling was used for modelling.

Table 2

Tentative identification of significant metabolites in patients undergoing ketamine treatment

	Compound	RT (min)	Measured mass (Da)	Mass error (p.p.m.)	Molecular formula	FC (%) in NRs	*P* value	CV (%) for QCs	Biochemical category	Subcategory
1	Phenyllactic acid§	13.58	166.0624	–3.5	C9H10O3	62.7	0.049	5.58	Phenolic, benzoyl, and phenyl derivatives	–
2	Phenylvaleric acid	10.78	178.0988	–3.3	C11H14O2	70.6	0.015	3.53	Phenolic, benzoyl, and phenyl derivatives	–
3	LPC (16:1)§	16.93	493.3162	–1.3	C24H48NO7P	–21.2	0.048	5.51	Glycerophos-pholipids	Monoacyl-glycerophos-phocholines
4	Deoxytetradecasphingenine	21.81	227.2240	–4.0	C14H29NO	–38.0	0.022	1.53	Sphingolipids	Sphingoid base
5	Deoxytetradecasphinganine	11.27	229.2394	–5.1	C14H31NO	17.4	0.029	7.99	Sphingolipids	Sphingoid base
6	Decanamide	9.23	171.1613	–5.9	C10H21NO	–46.3	0.045	2.20	Fatty acyls	Primary amides
7	Pentadecatetraenal	5.65	218.1663	–3.5	C15H22O	–42.1	0.021	14.80	Fatty acyls	Fatty aldehydes
8	Dimethyldioxododecatrienal	24.45	234.1269	5.6	C14H18O3	33.8	0.058	9.12	Fatty acyls	Fatty aldehydes
9	Hexadienoic acid	8.82	112.0523	–1.1	C6H8O2	143.4	0.030	8.37	Fatty acyls	Unsaturated FAs
	Hexenedial				C6H8O2				Fatty acyls	Fatty aldehydes
	Oxohexenal				C6H8O2				Fatty acyls	Fatty aldehydes
	Hydroxy-hexadienal				C6H8O2				Fatty acyls	Fatty aldehydes
	Dihydrobenzenediol				C6H8O2				Phenolic, benzoyl, and phenyl derivatives	
10	Hexadienoic acid	9.22	112.0523	–1.1	C6H8O2	145.1	0.029	7.69	Fatty acyls	Unsaturated FAs
	Hexenedial				C6H8O2				Fatty acyls	Fatty aldehydes
	Oxohexenal				C6H8O2				Fatty acyls	Fatty aldehydes
	Hydroxy-hexadienal				C6H8O2				Fatty acyls	Fatty aldehydes
	Dihydrobenzenediol				C6H8O2				Phenolic, benzoyl, and phenyl derivatives	–
11	Unknown§	7.35	162.1038		C11H14O	53.9	0.024	4.47		–

Tentative identification of metabolites significantly associated with response to ketamine treatment in 22 BD patients taking either lithium or VPA.
Note: For putative identification, all the compounds with a score >80% were first formula matched with the experimental isotopic pattern distribution on Mass Hunter software.
§Found in the comparison of Rs and NRs for lithium-subgroup in the same direction.
FC, fold change; FC was calculated as follows: (Average [NRs] – Average [Rs])/Average [Rs] × 100).
+/– , increase/decrease in NRs when compared with Rs.
LPC, lysophosphatidylcholine; this entity has been named with the number of carbon of the fatty acid attached to the backbone and the number of unsaturation, for example, LPC (16:1).
BD, bipolar depression; KET, ketamine; Li, lithium; NR, non-responders; R, responders; VPA, valproate.

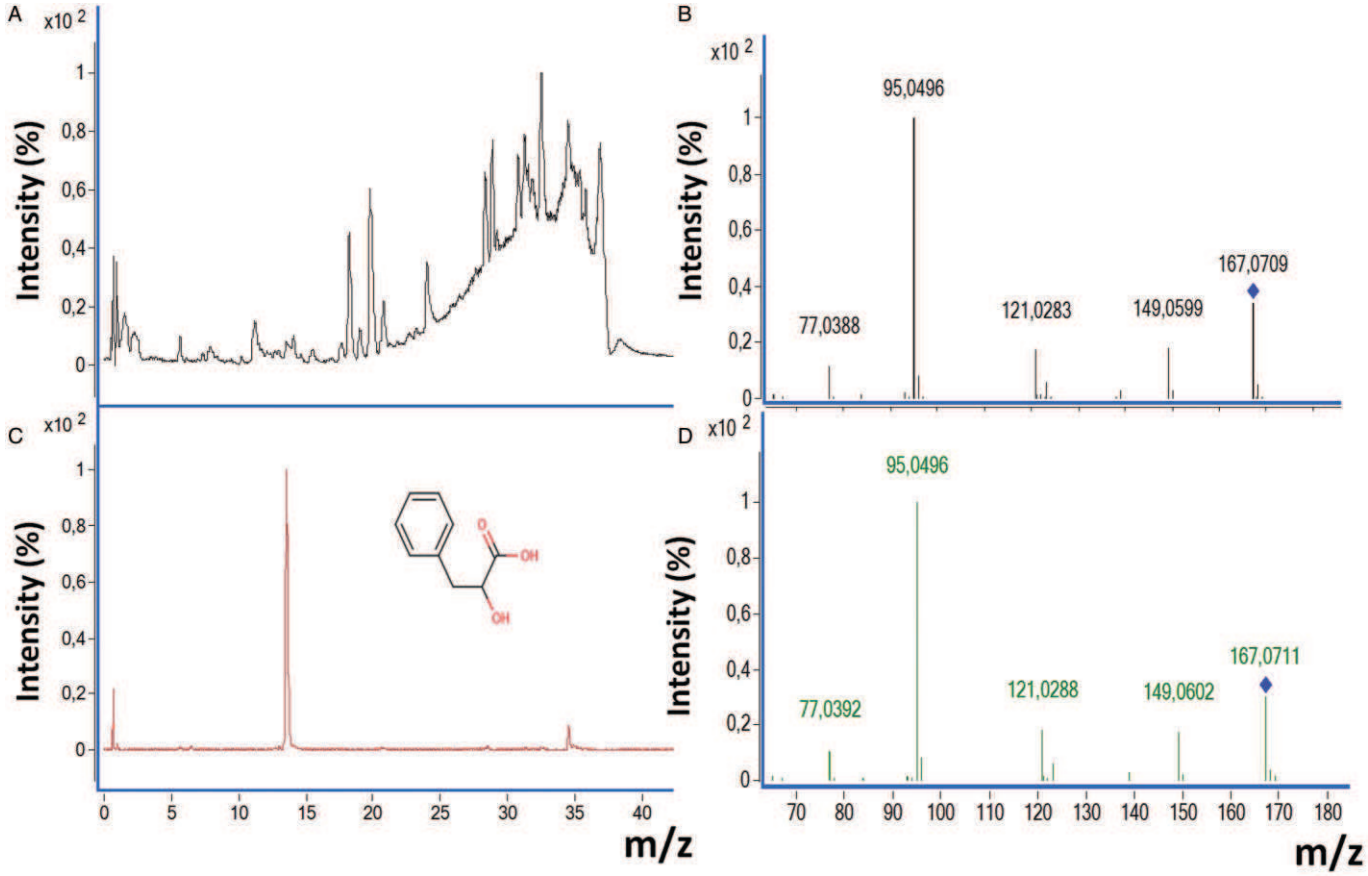


Figure 3

Identification of a selected biomarker using the example of PLA. (A) Total ion chromatogram from plasma sample of a BD patient was selected. (B) Extracted ion chromatogram of 167.0696 (m/z), retention time of 13.6 min. (C) MS/MS experimental spectrum from the ion [167.0696 (m/z), RT = 13.6 min]. (D) MS/MS spectrum of commercial analytical standard PLA reagent (collision energy = 175 V) was also compared with the experimental fragmentation.

response of patients with BD to treatment with ketamine. Thus, all of the patients were included in the analysis irrespective of the administered mood stabilizer. However, the presence of large differences in the pharmacometabolomic profiles associated with the valproate-subgroup and lithium-subgroup required an independent analysis of the two subgroups. Unfortunately, the number of patients in the valproate-subgroup (n = 6) was not large enough to achieve a significant statistical analysis of the data and only the data from the lithium-subgroup (n = 16) were re-analysed. Therefore, the chromatograms from the lithium-subgroup and QCs were re-aligned and filtered. A PLS-DA model was able to effectively (99.7%) classify the samples from the ketamine arm of the study with a prediction capacity of over 60% (Figure 2D). This indicated that there was an inherent metabolic change between Rs and NRs, which was not observed when the valproate-subgroup was included in the analysis. This model was subjected to an external cross-validation test that demonstrated that 87.5% of the samples were classified correctly.

The OPLS-DA model from Rs and NRs was used to select the statistically significant (jack-knife $P < 0.05$) biomarkers (Figure 2E) and 165 biomarker candidates were selected for further analysis. The identification process included total ion chromatogram, extracted ion chromatogram with retention time, MS/MS experimental spectrum and, when available, MS/MS spectrum of commercial analytical standard. An example of biomarker identification is presented using phenyl lactic acid (PLA), Figure 3. This process identified 18 compounds, 6 of which were increased and 12 were decreased in NRs relative to Rs. The identified metabolites are summarized in Table 3. Of these metabolites, lysophospholipids were confirmed through their characteristic fragments as described in the literature (Milne *et al.*, 2006). For lysophosphatidylcholines (LPCs), fragments 184.07, 104.11 and 86.1 m/z, and for lysophosphatidylethanolamines (LPEs), a fragment of ([M + H]-141.02 m/z) were observed. N_6,N_6,N_6-trimethyl-L-lysine (TML) and a majority of the LPCs (7/9) and LPEs (3/4) were decreased in NRs, relative to Rs, while monoglycerides (MG) and PLA were increased.

A good sample separation was obtained with plasma samples from the lithium-subgroup from the placebo arm of the study using the PLS-DA model (Figure 2F, Table 4). The comparison of the data from Rs and NRs in the placebo arm indicated that the same acyl-carnitines and glycerophospholipid compounds were significantly different and followed

Table 3

Metabolites significantly different between Rs and NRs taking lithium in the ketamine arm of the therapy

	Compound	RT (min)	Measured mass (Da)	Mass error (p.p.m.)	Molecular formula	MS/MS fragments	FC (%) in NRs	*P* value	CV (%) for QCs
1	Phenyllactic acid¶, §	13.58	166.0623	–4.2	C9H10O3	167.07, 149.06, 121.03, *95.05*, 77.04	47.9	0.028	5.1
2	N6,N6,N6-Trimethyl-L-lysine	0.61	188.1509	–10.1	C9H20N2O2	189.16, 173.02, 130.09, *84.08*	–22.9	JK	19.7
3	MG (16:0)	27.67	330.2765	–1.5	C19H38O4	331.28,313.27, 257.24, 239.24, 123.11, 109.10, 95.08, 85.10, 83.08, 81.07, 71.09, *57.07*	12.4	JK	12.1
4	MG (16:0)	28.38	330.2772	0.6	C19H38O4	331.28,313.27, 257.24, 239.24, 123.11, 109.10, 95.08, 85.10, 71.09, *57.07*	12.9	JK	7.9
5	LPE (18:2)	17.46	477.2854	–0.2	C23H44NO7P	478.30, *337.27*, 81.07	–39.1	JK	8.1
6	LPE (18:2)	18.12	477.2852	–0.7	C23H44NO7P	478.29, 338.28, *337.27*, 306.28, 135.11, 95.08, 81.07, 62.06	–46.1	0.029	3.2
7	LPE (18:1)	20.64	479.3006	–1.2	C23H46NO7P	480.31, 340.30, *339.29*, 308.29, 155.01, 62.06	–19.8	JK	2.8
8	LPE (22:6)	18.13	525.2848	–1.4	C27H44NO7P	526.28, *385.27*, 93.10, 79.06, 62.06	25.2	JK	3.3
9	LPC (16:1)§	16.94	493.3163	–1.1	C24H48NO7P	494.32, 476.31, *184.07*, 104.11, 86.10	–39.8	0.031	6.5
10	LPC (16:0)	19.06	495.3331	1.2	C24H50NO7P	496.33, *184.07*, 86.10	–14.2	JK	7.6
11	LPC (16:0)	19.87	495.3337	2.4	C24H50NO7P	496.34, *184.07*, 104.11, 86.10	–6.3	JK	4.1
12	LPC (18:2)	17.63	519.3325	0.0	C23H48NO7P	520.34, *184.07*, 86.10	–16.8	JK	11.2
13	LPC (18:2)	18.29	519.3334	1.8	C23H48NO7P	520.34, 502.33, *184.07*, 104.11, 86.10	–19.8	JK	10.8
14	LPC (20:3)	19.77	545.3479	–0.4	C28H52NO7P	546.35, *184.07*, 104.11, 86.10	–56.0	JK	16.7
15	LPC (20:2)	21.88	547.3622	–2.9	C28H54NO7P	548.37, *184.07*, 104.11, 86.10	–22.3	JK	17.9
16	LPC (22:6)	17.75	567.3311	–2.5	C30H50NO7P	568.34, *184.07*, 86.10	36.0	0.042	13.8
17	LPC (22:6)	18.27	567.3320	–0.9	C30H50NO7P	568.34, *184.07*, 104.11, 86.10	33.1	JK	19.2
18	Unknown§	7.35	162.1038	–	C11H15O	163.11, 135.08, *107.05*	41.5	0.026	5.1

Note: FC, fold change; FC was calculated as follows: (Average [NRs] – Average [Rs])/Average [Rs] × 100); +/– -increase/decrease in NRs compared with Rs; Unknown, compound which MS/MS spectra was not interpretable or not informative.
¶Confirmed also by standard
§Found in the comparison of Rs and NRs in combined subgroups (Li + VPA) in the same direction;
In MS/MS fragments the underlined number refers to the most abundant fragment observed.
MG, monoglyceride; LPE, lysophosphatidylethanolamine; LPC, lysophosphatidylcholine, these entities has been named with the number of carbon of the fatty acid attached to the backbone and the number of unsaturation, for example, LPC (16:1); JK, jack-knife analysis.
Li, lithium; NR, non-responders; R, responders; VPA, valproate.

Table 4
Tentative identification of metabolites significantly different between Rs and NRs taking lithium in the placebo arm of the therapy

	Compound	RT (min)	Measured mass (Da)	Mass error (p.p.m.)	Molecular formula	FC (%) in NRs	*P* value	CV (%) for QCs	Biochemical category	Subcategory
1[a]	Octanoylcarnitine	4.14	287.2090	−2.3	C15H29NO4	65.0	0.028	8.92	Fatty acyls	Fatty esters
2[a]	Decanoylcarnitine	8.18	315.2403	−2.1	C17H33NO4	66.6	0.024	7.64	Fatty acyls	Fatty esters
3	Dodecanoylcarnitine	12.10	343.2716	−1.9	C19H37NO4	40.7	0.049	5.59	Fatty acyls	Fatty esters
	N-palmitoyl serine				C19H37NO4				Fatty acyls	N-acyl amines
4[a]	LPC (16:1)	16.41	493.3166	−0.5	C24H48NO7P	−34.9	0.025	8.21	Glycerophospholipids	Monoacylglycerophosphocholines
	LPE (19:1)				C24H48NO7P				Glycerophospholipids	Monoacylglycerophosphoethanolamines
5	Trp Phe	1.03	351.1588	1.4	C20H21N3O3	−40.8	0.030	27.88	Dipeptide	

Note: For putative identification, all the compounds with a score >80% were first formula matched with the experimental isotopic pattern distribution on Mass Hunter software.
[a]Also found changed in the same direction in the comparison of Rs and NRs in the ketamine arm infusion.
FC, fold change; FC was calculated as follows: (Average [NRs] − Average [Rs])/Average [Rs] × 100).
+/−, increase/decrease in NRs when compared with Rs.
LPC, lysophosphatidylcholine; this entity has been named with the number of carbon of the fatty acid attached to the backbone and the number of unsaturation, for example, LPC (16:1).
KET, ketamine; NR, non-responders; R, responders.

the same pattern as observed in the combined (lithium + valproate) placebo group. The metabolites from placebo and ketamine arms did not match, suggesting that the metabolites identified in Table 3 are associated with response to ketamine therapy in the lithium-subgroup.

Discussion

Over the past five decades, a considerable number of antidepressants have been developed, representing a wide range of molecular and therapeutic classes. The primary pharmacological activities associated with these agents are either based on 5-HT (SSRI) or on noradrenaline (5-HT-noradrenaline reuptake inhibitors;SNRI). However, there are a substantial number of patients with major depression who do not respond to current antidepressant therapy and none of these agents have demonstrated a significant advantage in terms of clinical efficacy. In addition, there is a considerable latency in the onset of antidepressant effects, which typically take 6 weeks or more. Due to these limitations, new therapeutic targets are being explored, with the hopes of developing more effective and rapidly acting treatments. A primary focus in this effort is the development of glutamatergic modulators, particularly those acting at the NMDA receptor.

Ketamine, an NMDA receptor modulator, is currently under investigation for use in the treatment of depression and neuropathic pain. We have demonstrated that a subanesthetic dose of ketamine produces rapid antidepressant effects in patients diagnosed with BD (Diazgranados *et al.*, 2010) and MDD (Zarate *et al.*, 2006), and these effects can last up to 7 days (Zarate *et al.*, 2012). A simultaneous population pharmacokinetic model for ketamine and three of its major metabolites in BD patients has been developed, utilizing plasma samples collected up to 3 days after ketamine administration (Zhao *et al.*, 2012), and a pharmacodynamic study in a combined cohort of patients with MDD (n = 45) and BD (n = 22) has also been conducted (Zarate *et al.*, 2012). The data from these studies indicate that at 230 min after infusion, the clinical effect of ketamine robustly separates from placebo (Zarate *et al.*, 2006; 2012; Diazgranados *et al.*, 2010). In addition, 76% (25/33) of Rs at 230 min continued to meet response criteria at day 1 and only 10% (5/50) of NRs at 230 min were classified as Rs at a later time point. Based upon these data, the 230 min time point was chosen as the sampling point to capture clinical response and non-response. Thus, the preliminary metabolomic comparison of ketamine Rs and NRs in the treatment of BD was limited to the analysis of the 230 min plasma samples.

The initial analysis of 22 BD patients included in the study indicated that there were response-related differences in metabolomic patterns (Table 2), which was confirmed by PLS-DA and OPLS-DA models. After compound selection (Figure 2B) and database identification, 11 compounds were increased or decreased in Rs relative to NRs. A majority of the compounds were from the fatty acyl family. The observation of differences in fatty acyl metabolism between ketamine Rs and NRs was supported by analysis of plasma samples from the placebo arm in which acyl-carnitines were increased and L-carnitine decreased in NRs compared with Rs (Supporting Information Table S1).

While the differences between NRs and Rs were statistically significant in univariate analysis, in multivatiare analysis the differences were small and the quality of the models poor. A potential source of this result was the lack of homogeneity within the BD cohort produced by maintenance on either lithium or valproate. Therefore, the plasma samples from the ketamine arm of the study were divided into lithium – and valproate-subgroups based upon the mood stabilizers administered to the BD patients. The metabolomic pattern of each subgroup was determined and then compared in a Venn diagram, with the data from the analysis of the combined (lithium + valproate) group (Table 1). The patterns were significantly different with only two common compounds (Figure 2C). A similar result was obtained from the analysis of the samples from the placebo arm of the study, in which there was a greater than twofold increase in acyl-carnitines in the valproate-subgroup compared with lithium-subgroup (Supporting Information Table S3). This result is consistent with valproate-associated depletion of free carnitine due to formation of valproyl carnitine and decreased tubular reabsorption of acetylcarnitine (Lheureux and Hantson, 2009) and with the observations that valproate affects mitochondrial fatty acid oxidation (Silva *et al.*, 2008).

In the ketamine arm of the study, the relative concentrations of phenylalanine, tryptophan and bilirubin were increased in the lithium-subgroup and amino-octanoic acid, and arachidonoyl-serine was increased in the valproate-subgroup, independent of response to treatment with ketamine. Increased plasma levels of phenylalanine have been previously observed in heart failure patients diagnosed with MDD, relative to matched controls, suggesting a potential alteration in the phenylalanine metabolic pathway (Steffens *et al.*, 2010); and tyrosine plasma concentrations, a product of another metabolic pathway of phenylalanine, are also affected in the depressed patients (Nordin, 1988; Kaddurah-Daouk *et al.*, 2012). In addition, data from a proton nuclear magnetic resonance spectroscopy-based metabolonmic analysis using *post mortem* brain tissue from patients with a history of BD and rat brains obtained after chronic treatment with lithium or valproate indicate that valproate treatment increased glutamate levels and decreased glutamate/glutamine ratios, while lithium treatment increased GABA levels (Lan *et al.*, 2009). The results indicate that even though the concomitant administration of lithium or valproate does not produce a difference in the clinical response to ketamine therapy, it does produce a difference in the background metabolomic patterns. Based upon these differences, the analysis of the combined subgroups was not continued and only the plasma samples from the lithium-subgroup were subjected to additional analysis. It is important to note that these differences do not negate the use of a pretreatment metabolomics screen, but only highlights the challenges in application.

The metabolomic patterns in BD patients maintained on lithium were significantly different between Rs and NRs (Figure 2D and E). Eighteen of the 165 biomarker candidates selected for further analysis were identified, and the majority (15/18) of these were LPEs (4), LPCs (9) and MGs (2). The signals associated with 3/4 LPEs and 7/9 LPCs were increased in Rs relative to NRs, while the signals associated with both of the identified MGs were decreased in Rs (Table 2). The data indicate that there are differences in mitochondrial fatty acid metabolism between the two groups.

The analysis of the plasma samples also revealed that TML was increased in Rs, compared with levels in NRs. TML is a precursor of L-carnitine, a key factor in the mitochondrial β-oxidation of fatty acids. The differences in TML plasma level in Rs and NRs are consistent with the dissimilarities in fatty acid metabolism between the two groups. In addition, in the samples from the placebo arm of the study, octanoyl- and decanoyl-carnitines were up-regulated in NRs compared with Rs. While acyl-carnitines have antidepressant effects in elderly (Pettegrew *et al.*, 2000; 2002) and mice (Di Cesare Mannelli *et al.*, 2011), increased concentrations in NRs may reflect a dysfunction in signalling of the mitochondrial acyl-carnitine receptor. It is of interest to note that TML is synthesized from lysine by S-adenosyl-methionine (SAM) catalyzed N-methylation and that SAM is in turn synthesized from L-methionine. Higher levels of methionine were observed in MDD patients who were in remission (Kaddurah-Daouk *et al.*, 2012), which may reflect an alteration in the conversion of methionine to SAM affecting LPC synthesis.

In contrast to the increased levels of LPCs, LPEs and TML, the signals associated with PLA were decreased in Rs, compared with NRs. PLA is a product of the NADH-mediated reduction of phenylpyruvate and an indication of a disruption in phenylalanine metabolism. The potential alteration in phenylalanine metabolic pathways is in accord with previous observations that phenylalanine plasma levels are higher in heart failure patients diagnosed with MDD, relative to matched controls (Steffens *et al.*, 2010) and that the tyrosine pathway is compromised in depressed patients (Nordin, 1988; Kaddurah-Daouk *et al.*, 2012). Tyrosine is produced by the phenylalanine metabolic pathway. Moreover PLA and LPC (16:1) followed similar pattern of changes observed in plasma samples obtained from the combined group (lithium + valproate), suggesting that the differences in fatty acid metabolism and phenylalanine pathway were involved in the response or /non-response to ketamine, independent of the co-medication.

The results of this study indicate that in BD patients an underlying basis for a positive response or a non-response to treatment with ketamine is a difference in the mitochondrial metabolism of fatty acids. While the experimental approach used in this study, metabolomics, relies on an unbiased analysis of data, the identification of markers associated with mitochondrial function is not surprising. Mitochondrial function or dysfunction has been associated with mood disorders including BD (Horrobin and Bennett, 1999; Hroudova and Fisar, 2012; Tang and Wang, 2012). In a recent review of data from brain imaging studies and mitochondrial functional studies, the authors concluded that the results support the hypothesis of mitochondrial dysfunction in BD and suggest that BD is associated with decreased energy production and a shift towards anaerobic glycolysis (Minuzzi *et al.*, 2011). A key marker of changes in disease-related mitochondrial function is fatty acid and phospholipid metabolism (Horrobin and Bennett, 1999) and metabolomic studies have demonstrated that fatty acid metabolism is significantly lower and shifted from β-oxidation to ω-oxidation in depressed patients, compared with that in non-depressed controls (Maes *et al.*, 1996;

Paige *et al.*, 2007; Steffens *et al.*, 2010) and BD-related changes in phospholipid metabolism (Modica-Napolitano and Renshaw, 2004).

In summary, this is the first study of the plasma metabolomic patterns in patients receiving ketamine for the treatment of BD. The results indicate that there are distinct biochemical differences between patients who respond to treatment (Rs) and those who do not (NRs), and that the differences appear to be due to alterations in the mitochondrial metabolism of fatty acids. The major observation is that the differences in the metabolomics patterns observed between R and NRs were not produced by ketamine administration. Instead, they appear to set up the biochemical basis for the pharmacological response to ketamine. Thus, pretreatment metabolomics screening may be a guide to the prediction of response and a potential approach to the individualization of ketamine therapy in BD. In addition, these differences appear to be associated with disease-related dysregulation of mitochondrial function and networks. While the source of these differences is not clear, previous studies have identified genetic links between BD and variants in mitochondrial DNA (Hroudova and Fisar, 2012). Additional prospective studies will be required to better understand these observations.

Acknowledgements

This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIH) and the National Institute of Mental Health, NIH. The authors also acknowledge the funding from Spanish Ministerio de Economía y Competitividad (MEC) grant CTQ2011-23562 and A. V. acknowledges her fellowship to EADS CASA.

Conflict of interest

C. Z. and I. W. W. have submitted a patent for the use of ketamine metabolites in the treatment of bipolar disorder and major depression. They have assigned their rights in the patent to the U.S. government, but will share a percentage of any royalties that may be received by the government. A. R., M. S., G. L., C. B., M. P. L. report no potential conflicts of interest.

EADS CASA (who provided a Fellowship for A. V.) is an aeronautical company with no interest or property in the results.

References

Abo R, Hebbring S, Ji Y, Zhu H, Zeng Z-B, Batzler A *et al.* (2012). Merging pharmacometabolomics with pharmacogenomics using '1000 Genomes' single-nucleotide polymorphism imputation: selective serotonin reuptake inhibitor response pharmacogenomics. Pharmacogenet Genomics 22: 247–253.

Burroughs S, French D (2007). Depression and anxiety: role of mitochondria. Curr Anaesth Crit Care 18: 34–41.

Ciborowski M, Teul J, Martin-Ventura JL, Egido J, Barbas C (2012). Metabolomics with LC-QTOF-MS permits the prediction of disease stage in aortic abdominal aneurysm based on plasma metabolic fingerprint. PLoS ONE 7: e31982.

Di Cesare Mannelli L, Vivoli E, Salvicchi A, Schiovone N, Koverechi A, Messano M *et al.* (2011). Antidepressant-like effect of artemin in mice: a mechanism for acetyl-L-carnitine activity on depression. Psychopharmacology (Berl) 218: 347–356.

Diazgranados N, Ibrahim L, Brutsche NE, Newberg A, Kronstein P, Khalife S *et al.* (2010). A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. Arch Gen Psychiatry 67: 793–802.

Dolgin E (2013). Rapid antidepressant effects of ketamine ignite drug discovery. Nat Med 19: 8.

Horrobin DF, Bennett CN (1999). Depression and bipolar disorder: relationships to impaired fatty acid and phospholipid metabolism and to diabetes, cardiovascular disease, immunological abnormalities, cancer, ageing and osteoporosis. Possible candidate genes. Prostaglandins Leukot Essent Fatty Acids 60: 217–234.

Hroudova J, Fisar Z (2012). *In vitro* inhibition of mitochondrial respiratory rate by antidepressants. Toxicol Lett 213: 345–352.

Ji Y, Hebbring S, Zhu H, Jenkins GD, Biernacka J, Snyder K *et al.* (2011). Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. Clin Pharmacol Ther 89: 97–104.

Kaddurah-Daouk R, Krishnan K (2009). Metabolomics: a global biochemical approach to the study of central nervous system diseases. Neuropsychopharmacology 34: 173–186.

Kaddurah-Daouk R, Kristal BS, Weinshilboum RM (2008). Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol 48: 653–683.

Kaddurah-Daouk R, Yuan P, Boyle SH, Matson W, Wang Z, Zeng ZB *et al.* (2012). Cerebrospinal fluid metabolome in mood disorders-remission state has a unique metabolic profile. Sci Rep 2: 667.

Lan MJ, McLoughlin GA, Griffin JL, Tsang TM, Huang JTJ, Yuan P *et al.* (2009). Metabonomic analysis identifies molecular changes associated with the pathophysiology and drug treatment of bipolar disorder. Mol Psychiatry 14: 269–279.

Lheureux PE, Hantson P (2009). Carnitine in the treatment of valproic acid-induced toxicity. Clin Toxicol (Phila) 47: 101–111.

Lin L, Yu Q, Yan X, Hang W, Zheng J, Xing J *et al.* (2010). Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabonomics? A serum metabonomic study of kidney cancer. Analyst 135: 2970–2978.

Machado-Vieira R, Salvadore G, Luckenbaugh DA, Manji HK, Zarate CA Jr (2008). Rapid onset of antidepressant action: a new paradigm in the research and treatment of major depressive disorder. J Clin Psychiatry 69: 946 –958.

Maes M, Smith R, Christophe A, Cosyns P, Desnyder R, Meltzer HY (1996). Fatty acid composition in major depression: decreased ω3 fractions in cholesteryl esters and increased c20:4ω6/c20:5ω3 ratio in cholesteryl esters and phospholipids. J Affect Disord 38: 35–46.

Milne S, Ivanova P, Forrester J, Alex Brown H (2006). Lipidomics: an analysis of cellular lipids by ESI-MS. Methods 39: 92–103.

Minuzzi L, Behr GA, Moreira JCF, Frey BN (2011). Mitochondrial dysfunction in bipolar disorder: lessons from brain imaging and molecular markers. Rev Colomb Psiquiatr 40: 166S–182S.

Modica-Napolitano JS, Renshaw PF (2004). Ethanolamine and phosphoethanolamine inhibit mitochondrial function *in vitro*: implications for mitochondrial dysfunction hypothesis in depression and bipolar disorder. Biol Psychiatry 55: 273–277.

Nordin C (1988). Relationships between clinical symptoms and monoamine metabolite concentrations in biochemically defined subgroups of depressed patients. Acta Psychiatr Scand 78: 720–729.

Paige LA, Mitchell MW, Krishnan KR, Kaddurah-Daouk R, Steffens DC (2007). A preliminary metabolomics analysis of older adults with and without depression. Int J Geriatr Psychiatry 22: 418–423.

Pettegrew JW, Levine J, McClure RJ (2000). Acetyl-L-carnitine physical-chemical, metabolic, and therapeutic properties: relevance for its mode of *action* in Alzheimer's disease and geriatric depression. Mol Psychiatry 5: 616–632.

Pettegrew JW, Levine J, Gershon S, Stanley JA, Servan-Schreiber D, Panchalingam K *et al*. (2002). 31P-MRS study of acetyl-L-carnitine treatment in geriatric depression: preliminary results. Bipolar Disord 4: 61–66.

Quinones MP, Kaddurah-Daouk R (2009). Metabolomics tools for identifying biomarkers for neuropsychiatric diseases. Neurobiol Dis 35: 165–176.

Rubingh CM, Bijlsma S, Derks EPPA, Bobeldijk I, Verheij ER, Kachar S *et al*. (2006). Assessing the performance of statistical validation tools for megavariate metabolomics data. Metabolomics 2: 53–61.

Schwarz E, Bahn S (2008). The utility of biomarker discovery approaches for the detection of disease mechanisms in psychiatric disorders. Br J Pharmacol 153: S133–S136.

Silva MFB, Aires CCP, Luis PB, Ruiter JPN, Ijlst L, Duran M *et al*. (2008). Valproic acid metabolism and its effects on mitochondrial fatty acid oxidation: a review. J Inherit Metab Dis 31: 205–216.

Steffens DC, Jiang W, Krishnan KRR, Karoly ED, Mitchell MW, O'Connor CM *et al*. (2010). Metabolomic differences in heart failure patients with and without major depression. J Geriatr Psychiatry Neurol 23: 138–146.

Su YA, Wu J, Zhang L, Zhang Q, Su DM, He P *et al*. (2008). Dysregulated mitochondrial genes and networks with drug targets in postmortem brain of patients with posttraumatic stress disorder (PTSD) revealed by human mitochondria-focused cDNA microarrays. Int J Biol Sci 4: 223–235.

Tang V, Wang JF (2012). Oxidative stress in bipolar disorder. Biochem Anal Biochem S2–002. doi:10.4172/2161-1009.S2-002.

Zarate CA Jr, Singh JB, Carlson PJ, Brutsche ME, Ameli R, Luckenbaugh DA *et al*. (2006). A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. Arch Gen Psychiatry 63: 856–864.

Zarate CA Jr, Brutsche N, Laje G, Luckenbaugh DA, Venkata SLV, Ramamoorthy A *et al*. (2012). Relationship of ketamine's plasma metabolites with efficacy, diagnosis, and psychotomimetic effects in patients with treatment-resistant depression. Biol Psychiatry 72: 331–338.

Zhang L, Zhou R, Li X, Ursano RJ, Li H (2006). Stress-induced change of mitochondria membrane potential regulated by genomic and non-genomic GR signaling: a possible mechanism for hippocampus atrophy in PTSD. Med Hypotheses 66: 1205–1208.

Zhao X, Swarajya LV, Moaddel R, Luckenbaugh DA, Brutsche NE, Ibrahim L *et al*. (2012). Simultaneous population pharmacokinetic modeling of ketamine and three major metabolites in patients with treatment-resistant bipolar depression. Br J Clin Pharmacol 74: 304–314.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

http://dx.doi.org/10.1111/bph.12494

Figure S1 Statistical analysis of global profiles of patients taking lithium or valproate with ketamine /placebo. (A) PLS-DA and (B) OPLS-DA cross-validated plot of lithium (n = 10) versus valproate (n = 6) in ketamine arm of the therapy (Rs and NRs are combined within each group) [for (A), R^2Y = 0.99; Q^2 = 0.51 and for (B), R^2Y = 1.00; Q^2 = 0.80; n = 1080 features]. (C) PLS-DA and (D) OPLS-DA cross-validated plot of lithium (n = 7) versus valproate (n = 4) in placebo arm of the therapy (Rs and NRs were classified according to their response to ketamine therapy [for (A), R^2Y = 0.99; Q^2 = 0.85 and for (B), R^2Y = 1.00; Q^2 = 0.81; n = 1325 features]. Key: ■ – Li, ▲ – VPA; UV scaling was used for modelling.

Table S1 Tentative identification of metabolites in the placebo arm of the study after global profiling of the entire cohort.

Table S2 Tentative identification of metabolites significantly different between patients taking lithium or valproate in the ketamine arm of the therapy.

Table S3 Tentative identification of metabolites significantly different between patients taking lithium or valproate in the placebo arm of the therapy.

SUPPLEMENTARY INFORMATION.

Supplemental Data 1. Tentative identification of metabolites in the placebo arm of the study after global profiling of the entire cohort.

	Compound	RT (min)	Measured mass (Da)	Mass error (ppm)	Molecular Formula	FC[%] in NRs	*P* value	CV(%) for QCs	Biochemical category	Subcategory
1§	Octanoylcarnitine	4.14	**287.2090**	-2.3	C15H29NO4	51.4	0.042	8.92	Fatty Acyls	Fatty esters
2§	Decanoylcarnitine	8.18	**315.2403**	-2.1	C17H33NO4	59.5	0.038	7.64	Fatty Acyls	Fatty esters
3	Carnitine	0.73	**161.1040**	-7.4	C7H15NO3	-14.8	0.044	12.43	Fatty Acyls	Fatty esters
4§	LPC(16:1)	16.41	**493.3166**	-0.5	C24H48NO7P	-29.1	0.029	8.21	Glycerophospholipid	Monoacylglycero-phosphocholines
	LPE(19:1)				C24H48NO7P				Glycerophospholipid	Monoacylglycero-phosphoethanol-amines

NOTE: The metabolites were identified from plasma samples obtained from patients with BD (*n* =17) in the placebo arm. These patients were taking either Li or VPA as co-medication. For putative identification, all the compounds with a score >90% were first formula matched with the experimental isotopic pattern distribution on Mass Hunter software; FC - fold change; +/- - increased/decreased in NR group when compared to Rs; FC was calculated as follows: (Average [NRs]-Average [Rs])/Average [Rs]*100); § - metabolites found in lithium-subgroup in the same direction; LPE - lysophosphatidylethanolamine; LPC – lysophosphatidylcholine, these entities has been named with the number of carbon of the fatty acid attached to the backbone and the number of unsaturation, for example, LPC (16:1).

Supplemental Data 2. Tentative identification of metabolites significantly different between patients taking Li or VPA in the KET arm of the therapy.

	Compound	RT (min)	Measured mass (Da)	Mass error (ppm)	Molecular Formula	FC[%] in VPA	*P* value	CV(%) for QCs	Biochemical category	Subcategory
1	Phenylalanine	0.96	**165.0780**	–5.9	C9H11NO2	-70.65	0.033	3.27	Amino Acid	
2	Tryptophan	0.97	**204.0887**	–5.8	C11H12N2O2	-36.73	0.008	2.78	Amino Acid	
3	Tetradecenoylcarnitine	12.63	**369.2871**	–2.2	C21H39NO4	122.96	0.022	3.35	Fatty Acid Ester	Acyl Carnitine
4§	Tetradecenoylcarnitine	12.97	**369.2872**	–1.9	C21H39NO4	180.12	0.004	4.33	Fatty Acid Ester	Acyl Carnitine
5§	Tetradecanoylcarnitine	14.78	**371.3037**	0.4	C21H41NO4	93.68	0.017	4.46	Fatty Acid Ester	Acyl Carnitine
	N-stearoyl serine				C21H41NO4				Fatty Acyls	Fatty amide
6	Decanamide	11.28	**171.1615**	–4.8	C10H21NO	33.64	0.047	11.10	Fatty Acyls	Fatty amide
7	Arachidonoyl serine	11.30	**391.2719**	–0.9	C23H37NO4	212.01	0.005	2.67	Fatty Acyls	Fatty amide
	HETE-Ala				C23H37NO4				Fatty Acyls	Fatty amide
8§	Coenzyme Q1	12.20	**250.1197**	–3.2	C14H18O4	-59.85	0.000	2.43	Benzoquinones	
9	Amino octanoic acid	0.71	**159.1252**	–4.5	C8H17NO2	37.55	0.043	3.29	Fatty Acyls	Amino FAs
10	Deoxytetradecasphinganine	9.65	**229.2394**	–5.1	C14H31NO	-59.59	0.045	27.42	Sphingolipids	Sphingoid bases
11§	Oxo-dodecatrienoic acid	12.19	**208.1093**	–3.1	C12H16O3	-60.81	0.000	3.43	Fatty Acyls	Oxo fatty acids
12	Hydroxy-decadienediynoic acid	7.46	**180.0779**	–4.1	C10H12O3	-87.40	0.001	1.40	Fatty Acyls	Unsaturated fatty acids
	Methoxy-benzenepropanoic acid				C10H12O3				Phenolic, benzoyl, and phenyl derivatives	
13	Bilirubin / Bilirubin IXa	7.04	**584.2619**	–2.7	C33H36N4O6	-71.12	0.010	7.57	Heterocyclic Compounds	Bilirubin
	Dihydrobiliverdin				C33H36N4O6				Heterocyclic Compounds	Bilirubin
14§	Octadienoic acid	6.80	**140.0833**	–3.0	C8H12O2	VPA	NA	10.21	Fatty Acyls	Unsaturated FAs
	Octenedial				C8H12O2				Fatty Acyls	Fatty aldehydes
	Hexadienyl acetate				C8H12O2				Fatty Acyls	Fatty esters
	Oxo-octenal				C8H12O2				Fatty Acyls	Fatty aldehydes
	Propyl-pentadienoic acid				C8H12O2				Fatty Acyls	Branched FAs

15	Hydroxycorticosterone	4.30	**362.2084**	–2.5	C21H30O5	-43.82	0.014	2.68	Sterol Lipids	Steroids
	Dihydrocortisone				C21H30O5				Sterol Lipids	Steroids
	Trihydroxy-oxohexanorvitamin D3				C21H30O5				Sterol Lipids	Secosteroids
	Dihydroxy-oxopregnanal				C21H30O5				Sterol Lipids	Steroids
	Cortisol				C21H30O5				Sterol Lipids	Steroids
16§	Propyl-pentenoic acid-(Ene-VPA)	2.66	**142.0987**	–4.8	C8H14O2	VPA	NA	12.64	*Drug metabolite*	
17§	Valproic Acid	8.96	**144.1144**	–4.4	C8H16O2	VPA	NA	11.38	*Drug metabolite*	
18	Unknown	8.95	**485.2557**			VPA	NA	22.52		
19§	Unknown	8.95	**502.2828**			VPA	NA	NA		
20	Unknown	3.79	**537.1739**			VPA	NA	NA		
21§	Unknown	8.95	**629.3702**			VPA	NA	NA		
22§	Unknown	3.80	**200.0124**			VPA	NA	NA		
23§	Unknown	8.96	**239.0598**			VPA	NA	13.51		
24	Unknown	3.80	**245.0114**			VPA	NA	NA		
25§	Unknown	8.96	**256.0629**			VPA	NA	18.54		
26§	Unknown	8.48	**282.0426**			VPA	NA	NA		
27§	Unknown	8.95	**342.1471**			VPA	NA	17.64		
28	Unknown	8.95	**367.2024**			VPA	NA	NA		

NOTE: Tentative identification of metabolites that were significantly different between patients (n = 16) in the KET arm of the therapy taking either Lithium or VPA. For putative identification, all the compounds with a score >80% were first formula matched with the experimental isotopic pattern distribution on Mass Hunter software; FC - fold change; +/- - increased/decreased in valproate-subgroup when compared to lithium-subgroup; FC was calculated as follows: (Average [VPA]-Average[Li])/Average [Li]*100); compounds labeled as VPA means that they only appear in patients taking VPA; NA - not applicable, the mass is absent in one of the two groups or was not present in the QC sample because

it was below limit of detection; Unknown compounds that are present only in VPA group could possibly be metabolites of VPA; § - metabolites found in the same comparison in placebo arm following same pattern direction; FA - fatty acid.

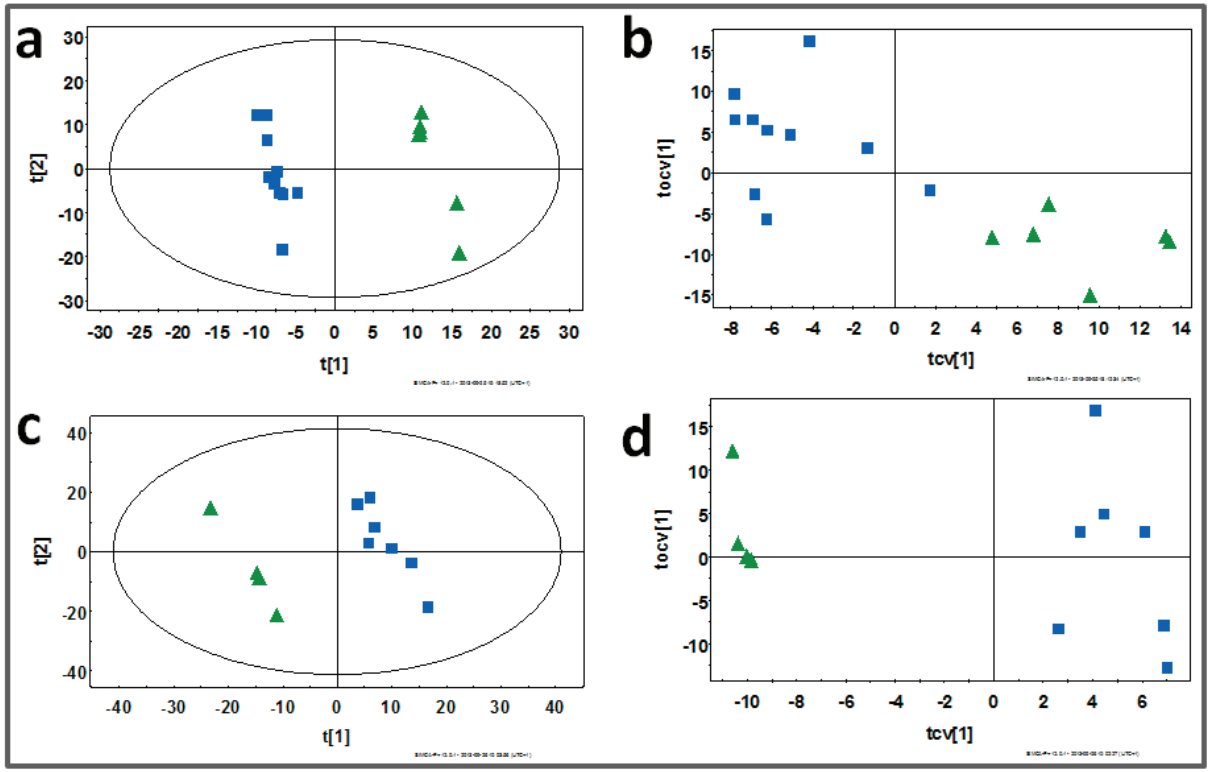
Supplemental Data 3. Tentative identification of metabolites significantly different between patients taking Li or VPA in the placebo arm of the therapy.

	Compound	RT (min)	Measured mass (Da)	Mass error (ppm)	Molecular Formula	FC[%] in VPA	p value	CV(%) for QCs	Biochemical category	Subcategory
1	Propionylcarnitine	0.72	**217.1300**	–6.5	C10H19NO4	157.05	0.099	7.90	Fatty Acyls	Acyl Carnitine
2	Butanoylcarnitine	0.94	**231.1459**	–4.9	C11H21NO4	151.94	0.038	16.01	Fatty Acyls	Acyl Carnitine
3	Dodecanoylcarnitine	12.10	**343.2716**	–1.9	C19H37NO4	100.68	0.016	5.59	Fatty Acyls	Acyl Carnitine
	N-palmitoyl serine				C19H37NO4				Fatty Acyls	Fatty amide
4§	Tetradecenoylcarnitine	13.64	**369.2887**	2.1	C21H39NO4	294.02	0.001	17.35	Fatty Acyls	Acyl Carnitine
5§	Tetradecanoylcarnitine	15.90	**371.3030**	–1.5	C21H41NO4	114.39	0.005	36.18	Fatty Acid Esters	Acyl Carnitine
	N-stearoyl serine				C21H41NO4				Fatty Acyls	Fatty amide
6	Octadecenoylcarnitine	20.21	**425.3497**	–1.9	C25H47NO4	50.75	0.027	18.49	Fatty Acyls	Acyl Carnitine
7§	Coenzyme Q1	11.58	**250.1200**	–2.0	C14H18O4	-55.33	0.018	8.06	Benzoquinone	
8§	Oxo-dodecatrienoic acid	11.58	**208.1092**	–3.6	C12H16O3	-57.19	0.014	9.63	Fatty Acyls	Oxo fatty acids
9	PGH2-EA (Prostaglandin H2-EA)	13.44	**393.2872**	–1.8	C23H39NO4	294.59	0.021	11.62	Fatty Acyls	Prostaglandins
10	Sphingosine	24.65	**299.2814**	–3.4	C18H37NO2	27.60	0.016	11.98	Sphingolipid	Sphingosines
	Ketosphinganine				C18H37NO2				Sphingolipid	Sphinganines
	Aminooctadecenediol				C18H37NO2				Fatty Acyls	Fatty alcohols
	Hydroxy-sphingosine				C18H37NO2				Sphingolipid	Sphingoid bases
	Sphingosine				C18H37NO2				Sphingolipid	Sphingoid bases
	Amino-octadecanoic acid				C18H37NO2				Fatty Acyls	Amino fatty acids
	Palmitoyl-EA				C18H37NO2				Fatty Acyls	Fatty amides
11§	Octadienoic acid	6.37	**140.0811**	–18.8	C8H12O2	VPA	VPA	16.41	Fatty Acyls	Unsaturated FAs
	Octenedial				C8H12O2				Fatty Acyls	Fatty aldehydes
	Hexadienyl acetate				C8H12O2				Fatty Acyls	Fatty esters
	Oxo-octenal				C8H12O2				Fatty Acyls	Fatty aldehydes
	Octynoic acid				C8H12O2				Fatty Acyls	Unsaturated FAs
	Propyl-pentadienoic acid				C8H12O2				Fatty Acyls	Branched FAs

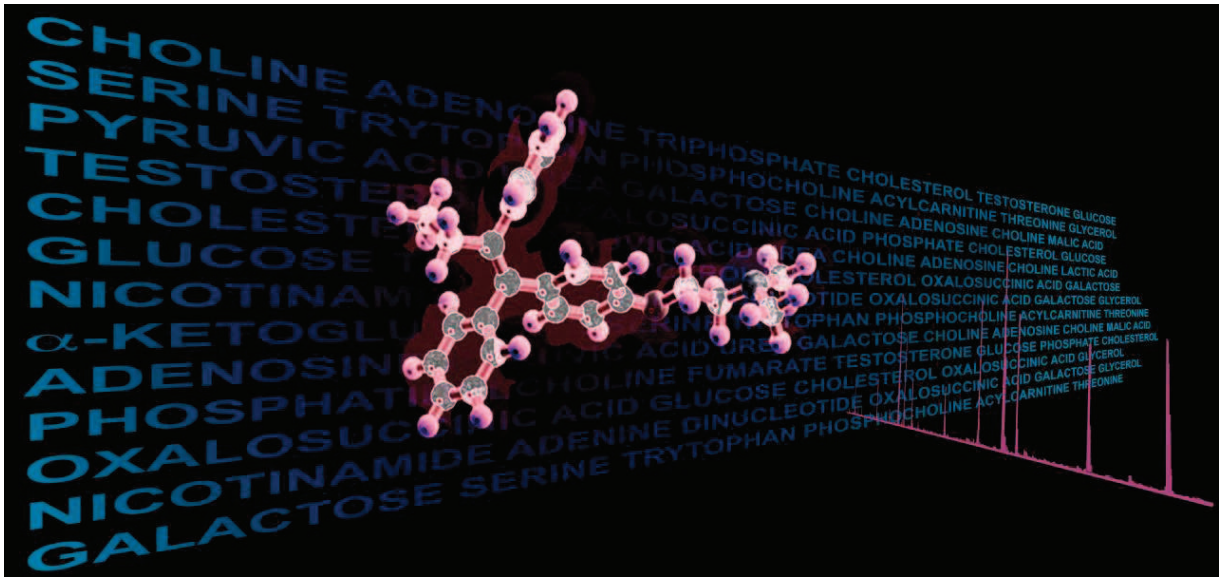
	Ethyl-hexenoic acid	7.99	**142.0988**	–4.0	C8H14O2	VPA	VPA	11.29	Fatty Acyls	Branched FAs
	Butyl-butenoic acid				C8H14O2				Fatty Acyls	Branched FAs
	Octenoic acid				C8H14O2				Fatty Acyls	Unsaturated FAs
12	Hydroxy-octenal				C8H14O2				Fatty Acyls	Fatty aldehydes
	Dimethyl-hexenoic acid				C8H14O2				Fatty Acyls	Branched FAs
	Methyl-heptenoic acid				C8H14O2				Fatty Acyls	Branched FAs
	Hexenyl acetate				C8H14O2				Fatty Acyls	Fatty ester
	Sphingosine	24.65	**299.2814**	–3.4	C18H37NO2	27.60	0.016	11.98	Sphingolipid	Sphingosine
	Ketosphinganine				C18H37NO2				Sphingolipid	Sphinganine
	Aminooctadecenediol				C18H37NO2				Fatty Acyls	Fatty alcohol
13	Hydroxy-sphingosine				C18H37NO2				Sphingolipid	Sphingoid base
	Sphingosine				C18H37NO2				Sphingolipid	Sphingoid base
	Amino-octadecanoic acid				C18H37NO2				Fatty Acyls	Amino fatty acid
	Palmitoyl-EA				C18H37NO2				Fatty Acyls	Fatty amide
14§	Valproic Acid	8.44	**144.1145**	–3.7	C8H16O2	VPA	VPA	9.19	*Drug metabolite*	
15§	Propyl-pentenoic acid-(Ene-VPA)	1.74	**142.0986**	–5.5	C8H14O2	VPA	VPA	VPA	*Drug metabolite*	
16§	Unknown	8.45	**239.0606**			VPA	VPA	NA		
17§	Unknown	7.99	**282.0462**			VPA	VPA	7.90		
18§	Unknown	8.44	**342.1482**			VPA	VPA	NA		
19§	Unknown	8.44	**629.3700**			VPA	VPA	NA		
20§	Unknown	8.45	**256.0639**			VPA	VPA	6.76		
21§	Unknown	3.53	**200.0125**			VPA	VPA	NA		
22§	Unknown	8.44	**502.2763**			VPA	VPA	NA		
23	Unknown	3.53	**169.9674**			VPA	VPA	NA		
24	Unknown	0.71	**185.1054**			VPA	VPA	NA		
25	Unknown	3.53	**212.0141**			VPA	VPA	NA		
26	Unknown	6.36	**235.0298**			VPA	VPA	NA		
27	Unknown	3.53	**258.0220**			VPA	VPA	12.08		
28	Unknown	4.04	**342.1287**			VPA	VPA	NA		
29	Unknown	8.44	**342.2680**			VPA	VPA	4.90		

30	Unknown	4.04	**374.0657**	VPA	VPA	6.42
31	Unknown	8.44	**387.1808**	VPA	VPA	NA
32	Unknown	5.34	**611.3290**	VPA	VPA	10.87

NOTE: For putative identification, all the compounds with a score above 75% were first formula matched with the experimental isotopic pattern distribution on Mass Hunter software; FC- fold change; +/- - increased/decreased abundance in valproate-subgroup when compared to lithium-subgroup; FC was calculated as follows: (Average [VPA]-Average[Li])/Average [Li]*100); compounds labeled as VPA means that they only appear in VPA group; NA - not applicable, the mass is absent in one of the two groups or it was not present in the QC sample because was below limit of detection; Unknown compounds that are present only in VPA group could be metabolites of VPA; § - also found in the comparison of Li versus VPA in the KET arm of therapy in the same direction; FA - fatty acid.



Supplemental Data 4. Statistical analysis of global profiles of patients taking lithium or VPA with KET/placebo. (A) PLS-DA and (B) OPLS-DA cross validated plot of Li (n=10) versus VPA (n=6) in KET arm of the therapy (Rs and NRs are combined within each group) [for (A), $R^2Y = 0.99$; $Q^2 = 0.51$ and for (B), $R^2Y = 1.00$; $Q^2 = 0.80$; n=1080 features]. (C) PLS-DA and (D) OPLS-DA cross validated plot of Li (n=7) versus VPA (n=4) in placebo arm of the therapy (Rs and NRs were classified according to their response to KET therapy [for (A), $R^2Y = 0.99$; $Q^2 = 0.85$ and for (B), $R^2Y = 1.00$; $Q^2 = 0.81$; n=1325 features]. Key: ■ - Li, ▲ - VPA; UV scaling was used for modeling.



THIRD SECTION.

TARGET ANALYSIS APPLICATION.

CHAPTER 6

3. Capillary Electrophoresis–with laser induced fluorescence detection (CE–LIF)

3.1. Fundamentals of the technique

Capillary electrophoresis (CE) fundamentals were discussed in Chapter 4. Summing this up, it is a powerful tool that represents a high resolution separation method which is especially suited for aqueous samples, enabling highly efficient separations (up to 10^6 theoretical plates[1-2]) of considerable number of sample components present in small sample volumes. Among other aspects, CE can be used with different types of detectors, of which the laser induced fluorescence detector (CE–LIF) is a very important one. In this section, only the characteristics of the LIF detector are going to be presented.

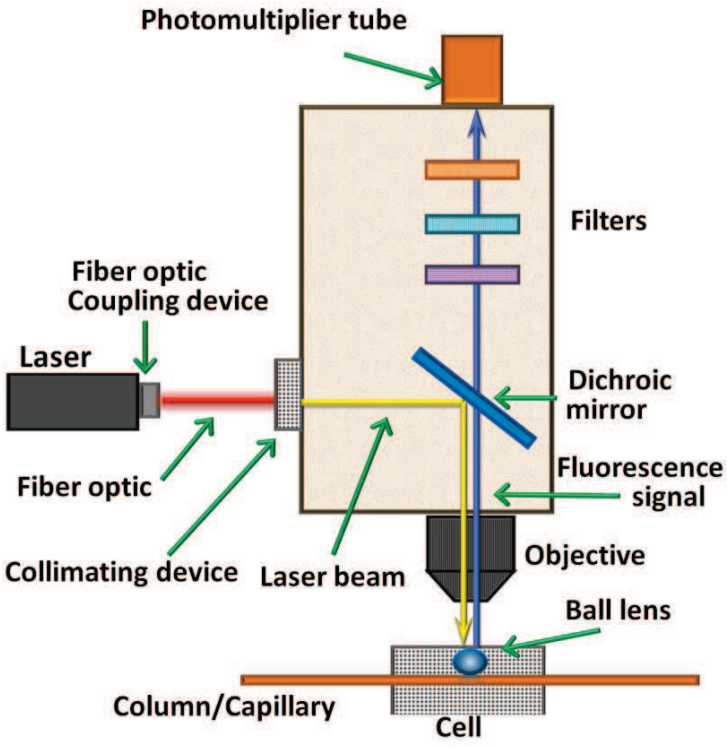


Figure 6.1. Scheme of a LIF detector, adapted from García–Campaña, A., 2007[1].

While CE has tremendous advantages as a separation technique, its main limitation is poor concentration detection limits based on the limited amount of sample volume that can be introduced into the capillary, and the short optical path–length. This is particularly observed with the most widely used detector for CE: UV absorbance[3–5]. Although UV absorbance is an almost universal detection system, it does not permit the analysis of compounds with low concentrations[1]. Thus, as an alternative method to improve sensitivity, the LIF detector was introduced.

LIF detectors use a laser to promote the analyte molecule to its excited electronic state at an appropriate wavelength. The resulting fluorescence emission is collected by suitable optics and focused onto an appropriate detector. A schematic representation of the components of the LIF detector is presented on Figure 6.1. A laser beam emits intense highly collimated light, which is adequate for focusing the light onto the small diameter of the capillary through the use of a microscope objective in conjunction with a ball lens. The emitted light is transmitted through the dichroic mirror to the emission filter and is detected by a photomultiplier. This has made LIF one of the most sensitive detection modes in CE, as compared with arc lamps used in conventional fluorescence, where the signal was lower by four to six orders of magnitude[1, 4]. In the case of some commercial brands, characteristics of each LIF detector are discussed elsewhere[4].

The fluorescence process results in exceedingly low limits of detection because this process is rapid enough that a single fluorophore can cycle repeatedly between excitation and emission before either exiting the detection window or undergoing photo–destruction. In addition, the emission spectrum is shifted relative to the excitation wavelength(s), so that the use of appropriate pass filters minimizes the amount of scattered excitation radiation detected[6]. Concentration detection limits under the best conditions are on the order of 100fM[7], although it has been reported to reach attomole or zeptomole levels and even a single molecule has been detected[5].

Strictly speaking, the fluorescence is observed only in compounds that contain an aromatic ring structure or some other series of conjugated *pi* bonds[6]. So in order to detect any analyte by LIF, this must present fluorescence. The lack of natural native fluorescence of most molecules implies a need to tag them by covalently bonding with fluorescent dyes; there is an extensive selection of reagents available to render most analytes of biological significance fluorescent[1, 4]. The majority of them react with primary and/or secondary amino groups of the analytes.

The ideal and universal fluorescent tag does not exist. The derivatization itself raises a wide range of problems especially in case of biological samples. For example, selectivity may be impaired since biofluids contain other compounds that are labelled along with the analytes of interest, noise may also be increased. Interfering peaks also derivated from the labelling reagent either as decomposition products or as excess of the reagent itself could be observed. Of note, for certain types of compounds, the use of

labelling may be impossible or have significant drawbacks. In complex structures such as proteins the formation of multiple analytes labelled could be possible, making their quantification difficult[5, 8].

Sensitivity and excitation selectivity of LIF detection can be controlled within limitations of practicality by employing different lasers; e.g. the Argon–ion laser (488 and 514nm) is the most common, with helium–cadmium laser (325 and 442nm) and blue diode laser (420nm)[1, 5].

3.2. Characteristics of CE–LIF for target analysis

Since the first introduction of the LIF detector for CE by Gassmann, E. in 1985[9], the limit of detection has improved dramatically compared to traditional UV detectors[8, 10], and therefore the instrumental and methodological developments for that range of applications for CE–LIF were increased. CE–LIF provided high speed, sensitivity and resolving power in the analyses of wide number components[3, 11]. The two main advantages are sensitivity and selectivity. Indeed[6], the LIF detector had been considered as highly selective based on its specific potential to detect native fluorescence of compounds of interest, and nearly universal because those analytes with no fluorophore, may be labelled (sample derivatization)[6].

The main drawback of CE–LIF is sample derivatization, because it frequently requires pre–column or post–column derivatization of the compounds[1]. Consequently, compounds susceptible to derivatization can be detected with LIF, making the process selective. Additionally, the use of LIF has the disadvantage that commercially available lasers have only a limited number of wavelengths[12]. This further reduces the detection applicability of LIF to certain types of compounds. For these reasons, CE–LIF is not an appropriate technique for large–scale metabolite analysis (or for non–target analysis)[13]. Instead, it has demonstrated that it can be very useful for targeted metabolomic studies[13–14]. The relative strengths and drawbacks of CE–LIF are summarized in Table 6.1. Of note: the separation characteristics of CE are the same as CE–UV (Chapter 4).

Table 6.1 Relative strengths and weaknesses of CE–LIF for metabolomics studies[13–14].

Strengths	Drawbacks
▪ High sensitivity and selectivity	▪ Not universal detector (only compounds native fluorescence, no suitable for non–target)
▪ High resolution, peak capacity& high efficiency	▪ Difficult to assign identification (standard spike)
▪ Selective through multiple separation modes (lower possibilities than CE–UV)	▪ Migration time shifting & problems with reproducibility
▪ Low consumption of organic solvent & reagents if needed	▪ Compounds with no native fluorescence sample must go under derivatization with the consequence of sample treatment
▪ Small amount of sample (few nanolitres)	▪ Detection is limited by the commercially laser wavelengths available and the restricted number of fluorophore
▪ Low cost& simple instrumentation	
▪ Suitable for chiral separations	

Based on the capabilities of CE–LIF, and its high sensitivity and selectivity, it has been used extensively in the analysis of peptides and proteins[8], of drugs in the pharmaceutical industry[4], for detecting important changes in biomolecules, and for diagnostic applications[1, 3, 10].

3.3. Insight into the applications of CE–LIF

Targeted analysis is a response to cover different aspects of research such as the validation of biomarkers. In other cases such as the next paper the analysis is focused on previous background knowledge targeting the metabolites of interest.

This study is based on treatment-resistant patients with bipolar disorder (BD) from previous chapter. BD among other neuropsychiatric diseases has shown differences in the levels of plasma amino

acids (normally when compared to healthy controls) and these have been associated with the symptomatology and severity of the disease[15–16]. Examples of amino acids for the disease are mentioned in the introduction to this paper, but it is important to highlight for this particular project that targeting these amino acids is a fundamental and complementary strategy alongside the non–target analysis from previous chapter.

The most suitable analytical method for target analysis will always rely on the characteristics of the metabolites of interest. In addition, it is important to consider the instrumental and economical resources available to perform this task as best as possible.

Over time, specific analytical methods for amino acids and their determination in biological samples have been developed. Extensive work in the use of different techniques has been applied, in order to obtain several amino acids in different kind of samples[14]. Amino acids are organic compounds that still represent a challenge in their quantification. Among the reasons are: (*i*) amino acids do not have a chromophore, so they do not absorb UV light[12]; (*ii*) most of them are highly hydrophilic, and therefore are difficult to extract using organic solvents for GC, or show poor retention in RP–LC and are difficult to separate from the solvent peak; and (*iii*) when they have gone through derivatization, it frequently results in more than one derivative per analyte, and some of these are lost during sample treatment. In addition, the concentration ranges are not the same for all amino acids; some are only present in trace levels. So along with the metabolite characteristics, the target methodology must be sensitive and selective enough for the metabolites of interest. Therefore, CE is a well–known and extensively used tool in the study of amino acids because CE offers high resolution valid for enantiomer separation, and in combination with LIF, it provides high sensitivity and selectivity with only small amount of sample[17]. Also, researchers are motivated by the great importance of these amino acids for clinical diagnosis in diseases, and by the rapidity and low cost of CE analysis. Some reviews[14, 18] describe the extensive advances in amino acid analysis by CE in all fields of applications.

Table 6.2 shows some examples of the different analytical techniques that have been employed in the analysis of amino acids in plasma. From this, different separation techniques such as LC (RP and ion exchange mode), GC and CE, and detection methods such as UV, LIF, MS (Q, QqQ, Q–IT) can be seen. Most of the methods use derivatization of the sample and lacks amino acid enantiomer determination. In

addition, for labelling reagents some of the main drawbacks are: (*i*) most of them (like NDA, CBQCA or OPA) use toxic chemicals such as sodium cyanide; (*ii*) or need an expensive UV laser; (*iii*) they can generate a large number of fluorescein related labelled by–products, which in some cases are not soluble in aqueous buffers; (*iv*) and most of them only react with primary amines.

As an advantage over these established methods, we propose the optimization of a previous method by using NBD–F (4 –Fluoro–7–nitro–2,1,3–benzoxadiazole) as labelling reagent, as this can be used for primary and secondary amines, and it can be detected by argon ion laser. Also, it requires a short reaction time and generates a low number of by–products. Along with the use of NBD–F for derivatization, the use of ultrafiltration devices (from Millipore, see conditions inside the article) allows the plasma to be deproteinized leading the sample, under derivatization, to give a very clear electropherogram.

Table 6.2. Examples of analytical methodologies and techniques used in the determination of amino acids in plasma samples.

Technique	Detector	Method characteristics	Sample preparation	Derivatization	Amino acids detected
RP–HPLC (C18)[19]	**UV and fluorescence detector**	Column: AccQ–Tag (150 X 3.9mm, 4µm) for amino acid hydrolysates (Waters, Milford, MA, USA). Detectors: photodiode array (UV at 248nm) and fluorescence (recorded at 395nm after excitation at 250nm).	Plasma was separated by centrifugation (1,000rpm, 10min, 10°C). For analysis plasma was diluted with water (1:5); 200µL were placed in a 10kD cutoff cartridges. Before use ultrafiltration membrane was pretreated with 200µL of 0.1M HCl and deionized water at 8,000rpm, 20min, 15°C. For diluted plasma 60min of same conditions as before were applied.	6–Aminoquinolyl–N–hydroxysuccinimidyl carbamate (AQC) solution (10mm) was prepared in ACN. 50µL of sample solution (ultrafiltrate) was mixed with 80µL of borate buffer plus 20µL of the IS solution and 350µL of water. Finally 20µL of AQC solution were added and vortex. Finally, a heating step (55°C, 10min) was applied.	26 amino acids detected LOD=1.5–15.4pmol/ LOQ=4.4–18.9pmol 75min of analysis No enantiomers
RP–HPLC (C18)[20]	**Fluorescence detector**	Column: XTerra column (100 X 4.6mm, 5µm) Waters (Milford, MA) Detector: Fluorescence detector (Waters 474) excitation at 340nm and an emission of 455nm	Plasma samples (20µL) were prepared by adding 20µL of an IS and 160µL of water. Proteins were precipitated with 200µL of 0.5M$HClO_4$. Samples were vortex and centrifuged (15,000xg, 5min, room temperature). 150µL of supernatant was filtered in a 0.2µm micro–centrifuge filter (15,000x*g* for 1min). 50µL of supernatant was put into a sample vial.	10 mg of o–phthalaldehyde (OPA) were dissolved in 0.2mL of methanol and 1.8mL of 200 mM tetraborate buffer (pH9.5) and 10µL of 3–mercaptopropionic acid (3–MPA) were added. This dissolution was diluted with 8mL of buffer. Finally 50µL of the labelling reagent were added to 50µL of the sample and incubated for 30 minutes before injection.	21 amino acids detected 35min of analysis No enantiomers

RP–HPLC (C18) LC X LC[21]	**UV and fluorescence detector**	Column: two Gemini–NX analytical columns (150 × 4.6mm, 3µm) connected in series. Detectors: UV–vis and a fluorescence detector.	Plasma samples were centrifuged (10,000×*g*, 1min) before 50µL of supernatant was mixed with 50µL IS. 200µL of cold ethanol was added. The mixture was centrifuged (10,000×*g*, 2min).	15µL of the supernatant was added to 65µL of borate buffer (AccQ–FluorTMkit) in a vial insert, and reacted with 20µLAccQ–Fluor™ reagent by immediate vortexfor 10 s.	33 amino acids detected LOD=0.1–8.0 ($X10^6$) pM/ LOQ=0.5–24.4 ($X10^6$) pM No enantiomers
HPLC (C8)[22]	**ESI–Q–MS**	Column: Wakosil–II 3C8–100HG column (100 × 2mm, 3µm) Detector: quadrupole(Q) with an ESI interface, operated in positive mode	Plasma was prepared by centrifugation (3,000rpm, 4°C, 15min). The samples were deproteinized using ACN at a final concentration of 80%. After centrifugation (15,000rpm, 4°C, 10min) supernatant was used for the analysis.	The amino acid concentrations in the plasma were measured by followed by precolumn derivatization with APDS (mix of N,N′–dihydroxysuccinimidyl carbonate and 3–aminopyridine), all the derivatization steps were performed by the autosampler	23 amino acids detected LOD=0.04–2.3 ($X10^3$)pmol 8min of analysis No enantiomers
GC[23]	**CI–QqQ–MS (MS/MS tandem detection)**	Column: A J&W Scientific DB–5MS column (30 m X 0.25mm, 0.25µm film thickness). Detector: Methane reagent gas was used for positive chemical ionization.	100µL serum were diluted with 4 volumes of methanol, kept on ice for 30min, centrifuged 10min at maximum speed, supernatant transferred to a new tube and vacuum–dried, and finally reconstituted directly with 1 M NaOH used in the derivatization protocol.	200µL of sample treated was mixed with 167µL of methanol and 34µL of pyridine. 20µL of methyl chloroformate (MCF) were added, plus 20µL of MCF followed by vortex. To separate the MCF derivatives 400µL of chloroform and 400µL of Na–bicarbonate solution were added[24].	23 amino acids detected LOD=0.3–1.1 ($X10^{-6}$)pM 15.5min of total run time No enantiomers

Amino Acid Analyzer[25]	**Spectro–photometric**	Cation–exchange chromatography	Frozen plasma samples were thawed, mixed with 5–sulfosalicylic acid (final conc. 2.3%), then centrifuged (10min, 11,000x*g*, 4°C).	Post–column reaction with ninhydrin reagent	33 amino acids detected No enantiomers
CE (MEKC)[26]	**LIF**	Capillary: bare fused–silica capillary (50cm to the detector, 75µm i.d.) Detector: LIF with argon ion laser (3mW), excitation at 488 nm and recorded at 520nm.	Blood was centrifuged (10 000rpm, 10min), 100µL of sample was deproteinized by mixing with 200µL of ACN. After vortex (2min), the mixture was centrifuged (10 000rpm, 10min). Supernatant was used for derivatization.	80µL of 15mM of Br–BQCA was added to the appropriate sample volume, plus 40µL of cyanide solution. The mixture was incubated (55°C, 50min) in the dark. Before analysis, mix was diluted to 1mL with the mixture of methanol and running buffer.	18 amino acids detected LOD=0.6–73 ($X10^3$) pmol 33min of analysis No enantiomers
CE (CZE)[27]	**C^4D**	Capillary: uncoated fused silica capillary (25mm id, 33cm in total length) Detector: C^4D (contactless conductivity detection).	50µL of plasma was deproteinized by adding 100µL ACN. After shaking, the mixture was centrifuged (3min, 4000x*g*). Supernatant was then filtered through a Durapore membrane (pore size 0.45mm) and then analysed.	No derivatization process	25 amino acids detected LOD=0.1–1.3 ($X10^6$) pmol 50min of analysis No enantiomers
CE (MEKC)[28]	**LED–IF (laser emitting diode–induced fluorescence)**	Capillary: fused–silica capillary (75µm i.d.) Detector: A violet LED luminous intensity of 300 mcd (operating current: 20 mA; peak emission: 410nm).	50µL of plasma was mix with 100µL NaCN (0.2mM), 100µL NDA (0.2mM), 50µL sodium tetraborate (10mM) and 200µL water. After gentle shaking, the solution was incubated at room temperature for 30 min.	Derivatization of DL–amino acids was done on a precolumn. Equal volumes (50µL) of 10mM $Na_2B_4O_7$, 100mM NaCN, and 25mM NDA solutions were mixed with 5µL of certain concentration of amino acids; the mixture was diluted to 500 mL with water.	3 pair of DL–amino acids LOD=7.1–57 ($X10^3$) pmol 30min of analysis

CE (CZE)[29]	**ESI–Q–IT MS (MS/MS tandem detection)**	Capillary: fused silica uncoated capillary (50µm id, total length of 79.5cm). Detector: Quadrupole– Ion Trap, ESI interface.	50 µL L of plasma was deproteinized by adding 15µL of formic acid (1% solution) and 100µL of cold ACN (4°C), after it was centrifuged (13 400rpm, 5min) and 15µL of the supernatant was used for injection.	No derivatization process	L–arginine and its methylated metabolites LOD=down to 10 ($X10^3$)pmol 12min per run

After the optimization and validation our method we were able to detect 14 amino acids in plasma and the enantiomers of all were satisfactorily separated (with a resolution of ≥ 1.5). This method was applied to BD patients that were taking lithium as a mood stabilizer. The results from the amino acid quantification showed concordant results with the literature for patients with a depressive or BD disorder.

As an additional advantage of this method is the possibility of measuring the chirality of amino acids. This ability is of growing importance in neuroscience where it is thought that particular D–amino acid forms may function as neuromodulators, and perhaps play an important role in early nervous system development[6]. Therefore, specifically for BD disease, D–serine plasma levels have been suggested as a biomarker for psychosis[30–31]. In addition, D–serine produces antidepressant–like effects in rodents and humans when used as a treatment. Here D–serine acts as a co–agonist of glutamate that binds with high affinity to the NMDAR unit which, unlike under ketamine, acts as a block to this receptor. The role of D–serine when ketamine is administrated has not yet been established, and was one of the objectives pursued with this methodology.

Although the data was unpublished, L– and D–serine levels were measured from patients with BD under the ketamine arm of the study, who were also using lithium or valproic acid as mood stabilizers (samples from clinical study in Chapter 5). The quantification and comparison between responders (Rs) and non–responders (NRs) is presented in Figure 6.2. Here the concentrations regardless of the mood stabilizer (Figure 6.2A) and within the lithium group only (Figure 6.2B) are presented. As can be observed in the combined cohort, D–serine concentrations were significantly higher (p=0.038) in NRs (mean 3.7μM ± 0.9) when compared to the Rs (mean 2.9μM ± 0.6). In patients taking only lithium, D–serine was significantly higher ((p=0.038) in NRs (mean 3.8μM ± 0.8) when compared to the Rs (mean 2.9μM ± 0.6). There was a trend for higher L–serine in NRs, but there were no significant differences. Therefore the quantification of plasma D–serine levels represents a potential biomarker of antidepressant response to ketamine infusion.

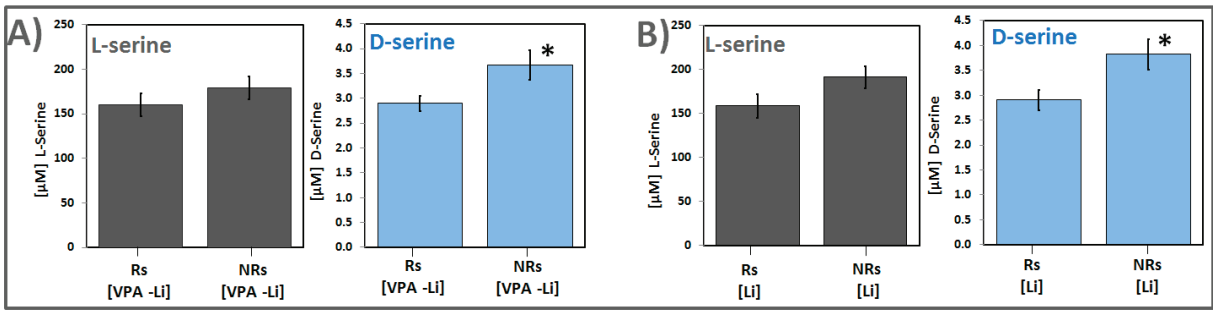


Figure 6.2. Quantification of D– and L–serine in plasma from BD patients. (A) Comparison total L– and D– serine concentrations in Rs (n=13) and NRs (n=9) regardless of co–medication at 230 min after KET administration; (B) L– and D– serine differences in Rs (n=9) and NRs (n=7) when only taking lithium for plasma after KET infusion. The results were analyzed by unpaired two–tailed Student's *t*–test. *Indicates $p<0.05$.

These findings were the key in confirming the same pattern in the complete set of plasma samples, in an additional set of samples taken from patients with major depressive disorder (MDD), who were under ketamine treatment (article under revision). It seems that the use of this analytical method and the quantification of D–serine as a biomarker for the prediction of the response to ketamine treatment is quite promising.

REFERENCES.

1. Garcia–Campana, A. M.; Taverna, M.; Fabre, H., LIF detection of peptides and proteins in CE. *Electrophoresis* **2007,***28* (1–2), 208–32.

2. Lapainis, T.; Rubakhin, S.; Sweedler, J., Capillary electrophoresis with electrospray ionization mass spectrometric detection for single–cell metabolomics. *Anal Chem* **2009,***81* (14), 5858–64.

3. Szöko, E.; Tábi, T., Analysis of biological samples by capillary electrophoresis with laser induced fluorescence detection. *J Pharm Biomed Anal* **2010,***53* (5), 1180–92.

4. Couderc, F.; Causse, E.; Bayle, C., Drug analysis by capillary electrophoresis and laser–induced fluorescence. *Electrophoresis* **1998,***19* (16–17), 2777–90.

5. Tseng, H. M.; Li, Y.; Barrett, D. A., Bioanalytical applications of capillary electrophoresis with laser–induced native fluorescence detection. *Bioanalysis* **2010,***2* (9), 1641–53.

6. Lapainis, T.; Sweedler, J. V., Contributions of capillary electrophoresis to neuroscience. *J Chromatogr A* **2008,***1184* (1–2), 144–58.

7. Johnson, M. E.; Landers, J. P., Fundamentals and practice for ultrasensitive laser–induced fluorescence detection in microanalytical systems. *Electrophoresis* **2004,***25* (21–22), 3513–27.

8. Lacroix, M.; Poinsot, V.; Fournier, C.; Couderc, F., Laser–induced fluorescence detection schemes for the analysis of proteins and peptides using capillary electrophoresis. *Electrophoresis* **2005,***26* (13), 2608–21.

9. Gassmann, E.; Kuo, J. E.; Zare, R. N., Electrokinetic separation of chiral compounds. *Science* **1985,***230* (4727), 813–4.

10. Ban, E.; Song, E. J., Recent developments and applications of capillary electrophoresis with laser–induced fluorescence detection in biological samples. *J Chromatogr B Analyt Technol Biomed Life Sci* **2013,***929*, 180–6.

11. Lin, Y. W.; Chiu, T. C.; Chang, H. T., Laser–induced fluorescence technique for DNA and proteins separated by capillary electrophoresis. *J Chromatogr B Analyt Technol Biomed Life Sci* **2003,***793* (1), 37–48.

12. Prata, C.; Bonnafous, P.; Fraysse, N.; Treilhou, M.; Poinsot, V.; Couderc, F., Recent advances in amino acid analysis by capillary electrophoresis. *Electrophoresis* **2001,***22* (19), 4129–38.

13. Ramautar, R.; Demirci, A.; de Jong, G. J., Capillary electrophoresis in metabolomics. *Trends in Analytical Chemistry* **2006,***25* (5), 455–466.

14. Poinsot, V.; Carpene, M. A.; Bouajila, J.; Gavard, P.; Feurer, B.; Couderc, F., Recent advances in amino acid analysis by capillary electrophoresis. *Electrophoresis* **2012,***33* (1), 14–35.

15. Hoekstra, R.; Fekkes, D.; Loonen, A. J.; Pepplinkhuizen, L.; Tuinier, S.; Verhoeven, W. M., Bipolar mania and plasma amino acids: increased levels of glycine. *Eur Neuropsychopharmacol* **2006,***16* (1), 71–7.

16. Mitani, H.; Shirayama, Y.; Yamada, T.; Maeda, K.; Ashby, C. R.; Kawahara, R., Correlation between plasma levels of glutamate, alanine and serine with severity of depression. *Prog Neuropsychopharmacol Biol Psychiatry* **2006,***30* (6), 1155–8.

17. Poinsot, V.; Bayle, C.; Couderc, F., Recent advances in amino acid analysis by capillary electrophoresis. *Electrophoresis* **2003,***24* (22–23), 4047–62.

18. Poinsot, V.; Rodat, A.; Gavard, P.; Feurer, B.; Couderc, F., Recent advances in amino acid analysis by CE. *Electrophoresis* **2008,***29* (1), 207–23.

19. Jaworska, M.; Stańczyk, M.; Wilk, M.; Kłaczkow, G.; Anuszewska, E.; Barzał, J.; Rzepecki, P., New approach for amino acid profiling in human plasma by selective fluorescence derivatization. *Amino Acids* **2012,***43* (4), 1653–61.

20. Frank, M. P.; Powers, R. W., Simple and rapid quantitative high–performance liquid chromatographic analysis of plasma amino acids. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007,***852* (1–2), 646–9.

21. Wang, H.; McNeil, Y. R.; Yeo, T. W.; Anstey, N. M., Simultaneous determination of multiple amino acids in plasma in critical illness by high performance liquid chromatography with ultraviolet and fluorescence detection. *J Chromatogr B Analyt Technol Biomed Life Sci* **2013,***940*, 53–8.

22. Shimbo, K.; Oonuki, T.; Yahashi, A.; Hirayama, K.; Miyano, H., Precolumn derivatization reagents for high–speed analysis of amines and amino acids in biological fluid using liquid chromatography/electrospray ionization tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2009,***23* (10), 1483–92.

23. Kvitvang, H. F.; Andreassen, T.; Adam, T.; Villas–Boas, S. G.; Bruheim, P., Highly sensitive GC/MS/MS method for quantitation of amino and nonamino organic acids. *Anal Chem* **2011,***83* (7), 2705–11.

24. Villas–Boas, S. G.; Delicado, D. G.; Akesson, M.; Nielsen, J., Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography–mass spectrometry. *Anal Biochem* **2003,***322* (1), 134–8.

25. Hisamatsu, T.; Okamoto, S.; Hashimoto, M.; Muramatsu, T.; Andou, A.; Uo, M.; Kitazume, M. T.; Matsuoka, K.; Yajima, T.; Inoue, N.; Kanai, T.; Ogata, H.; Iwao, Y.; Yamakado, M.; Sakai, R.; Ono, N.; Ando, T.; Suzuki, M.; Hibi, T., Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. *PLoS One* **2012,***7* (1), e31131.

26. Zhang, N.; Zhang, H. S.; Wang, H., Separation of free amino acids and catecholamines in human plasma and rabbit vitreous samples using a new fluorogenic reagent 3–(4–bromobenzoyl)–2–quinolinecarboxaldehyde with CE–LIF detection. *Electrophoresis* **2009,***30* (13), 2258–65.

27. Tuma, P.; Malkova, K.; Samcova, E.; Stulik, K., Rapid monitoring of arrays of amino acids in clinical samples using capillary electrophoresis with contactless conductivity detection. *J Sep Sci* **2010,***33* (16), 2394–401.

28. Tseng, W. L.; Hsu, C. Y.; Wu, T. H.; Huang, S. W.; Hsieh, M. M., Highly sensitive detection of chiral amino acids by CE based on on–line stacking techniques. *Electrophoresis* **2009,***30* (14), 2558–64.

29. Desiderio, C.; Rossetti, D. V.; Messana, I.; Giardina, B.; Castagnola, M., Analysis of arginine and methylated metabolites in human plasma by field amplified sample injection capillary electrophoresis tandem mass spectrometry. *Electrophoresis* **2010,***31* (11), 1894–902.

30. Hashimoto, K.; Fukushima, T.; Shimizu, E.; Komatsu, N.; Watanabe, H.; Shinoda, N.; Nakazato, M.; Kumakiri, C.; Okada, S.; Hasegawa, H.; Imai, K.; Iyo, M., Decreased serum levels of D–serine in patients with schizophrenia: evidence in support of the N–methyl–D–aspartate receptor hypofunction hypothesis of schizophrenia. *Arch Gen Psychiatry* **2003,***60* (6), 572–6.

31. Calcia, M. A.; Madeira, C.; Alheira, F. V.; Silva, T. C.; Tannos, F. M.; Vargas–Lopes, C.; Goldenstein, N.; Brasil, M. A.; Ferreira, S. T.; Panizzutti, R., Plasma levels of D–serine in Brazilian individuals with schizophrenia. *Schizophrenia research* **2012,***142* (1–3), 83–7.

Mª Paz Lorenzo[1]
Alma Villaseñor[1]
Anuradha Ramamoorthy[2]
Antonia Garcia[1]

[1]Center for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Campus Montepríncipe, Madrid, Spain
[2]Laboratory of Clinical Investigation, Intramural Research Program, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

Received November 23, 2012
Revised February 6, 2013
Accepted February 7, 2013

Research Article

Optimization and validation of a capillary electrophoresis laser-induced fluorescence method for amino acids determination in human plasma: Application to bipolar disorder study

Quantitative and qualitative analysis of amino acids in biofluids offers relevant information in diagnosis of diseases, evaluation of nutritional state, and in elucidating metabolic influences on physiology. A simple, rapid, and robust procedure in terms of sample treatment, separation, and quantitation based on CE-LIF has been optimized for use in human plasma samples. Time required for derivatization was 15 min and analysis time was 35 min. 4-Fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F) was the labeling agent used for obtaining fluorescent derivatives. Electrophoretic conditions were: 175 mM borate buffer at pH 10.25 prepared with 12.5 mM β-cyclodextrin. The voltage applied was +21 kV. Fourteen amino acids could be quantified: L-proline, L-phenylalanine, L-leucine, L-isoleucine, L-ornithine, D-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L-serine, D-serine, taurine and L-glutamate. With this chiral CE-LIF method, L- and D-amino acids are adequately separated. The method was validated for a representative group of amino acids in human plasma: L-proline, L-isoleucine, L-ornithine, L-glutamine, L-alanine L-threonine, glycine, L-serine, D-serine, and glutamate. The method has been successfully applied to human plasma from patients with bipolar disorder, all of whom were taking lithium as a mood stabilizer. Eleven amino acids were quantified in plasma from nine patients, aged 24–55 years. The results were in accordance to published values for the bipolar patients. The method is useful particularly in studies where plasma amino acid levels can be used as biomarkers for diagnosis of diseases, evaluating the disease progression, and monitoring response to drug therapy.

Keywords:
Amino acids / Biofluid / Bipolar disorder / Fluorescence / Plasma
DOI 10.1002/elps.201200632

1 Introduction

Amino acids are organic compounds that play a major role in a number of important biological processes including energy metabolism, neurotransmission, lipid transport, etc. Quantitative and qualitative analysis of amino acids has been applied in the diagnosis of diseases (e.g. inborn errors of metabolism), evaluation of nutritional state, and in elucidating metabolic influences on physiology [1–3]. Plasma levels of amino acids have been shown to be associated with symptoms and severity of a number of neuropsychiatric diseases such as schizophrenia, depression, and bipolar disorder (BD) [3, 4]. BD, formerly known as manic-depressive psychosis, is one of the most debilitating and common psychiatric disorders worldwide. BD is well distinguished by flashing emotional and behavioral disruptions [5]. As free amino acids are important for neurotransmission and receptor function, changes in their metabolism can be used not only for disease diagnosis, but also for monitoring treatment outcomes. Over or under expression of specific amino acids have been observed in patients of neuropsychiatric conditions when compared to controls: for example, some authors have found higher concentrations of glutamate, glutamine, and glycine in BD patients [4,6]; depressed patients had a significant increase in plasma levels of glutamate, glycine, glutamine, and taurine when compared to controls [3]. Studies have suggested that

Correspondence: Dr. Antonia García, Center for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Campus Montepríncipe, 28668 Boadilla del Monte (Madrid), Spain
E-mail: antogar@ceu.es
Fax: +34-91-3724712

Abbreviations: **BD**, bipolar disorder; **FQ**, 3-(2-furoyl) quinoline-2-carboxaldehyde; **IS**, internal standard (aminoadipic acid); **NBD-F**, 4-fluoro-7-nitro-2,1,3-benzoxadiazole; **NDA**, naphthalene dicarboxaldehyde; **OPA**, *O*-phthaldialdehyde

Colour Online: See the article online to view Figs. 1 and 2 in colour.

higher plasma serine concentration is a possible biomarker for schizophrenia, mania, paranoia, psychotic depression, and unipolar depression [7–9]. Thus, these neuroactive amino acids can be utilized as biomarkers in the diagnosis and prognosis of a number of neurological disorders. Some of these amino acids are present in trace levels in biological fluids. Increased complexity of the biological matrices makes it even more difficult to quantify these amino acids [10]. Thus, it is necessary to develop a highly sensitive analytical methodology for the determination of these amino acids. As was described elsewhere [11], quantitative measurement of the complete range of amino acids in biological samples is an important challenge in clinical biochemistry for several reasons: (i) amino acids do not have a chromophore; (ii) most of them are highly hydrophilic and therefore are difficult to extract using organic solvents for GC, show poor retention in RP-LC, and are difficult to separate from the solvent peak. The chromatographic separation using RP needs analyte derivatization or the use of ion-paring agents to increase the chromatographic retention of analytes and to preclude their co-elution with the void volume; (iii) in GC, the derivatized compounds are very volatile, usually resulting in more than one derivative per analyte and some of these are lost during the sample treatment. Although, GC-MS involves time-consuming derivatization or complicated extraction procedures, it is the gold standard technique for the diagnosis of inborn errors of metabolism.

Different analytical techniques based on chromatographic and electrophoretic methodologies have been reported in the literature for analyzing amino acids from different biological matrices. Among the LC techniques, both ion exchange, RP-HPLC, and RP-HPLC-MS are used extensively. Other common LC based methods include hydrophilic interaction chromatography [12]. Different detectors can been used with LC including UV [13, 14], fluorescence [15, 16], and MS [17]. With UV-VIS and fluorescence detection, derivatization (pre or post separation) is often necessary due to the lack of a chromophore [18, 19]. Ion exchange separation usually requires a postcolumn derivatization with ninhydrin [18]. *O*-phthaldialdehyde (OPA) in combination with 2-mercaptoethanol is used to produce fluorescent isoindole derivatives that can be separated by HPLC and detected fluorimetrically. However, this reagent reacts only with primary amines in the presence of thiol and generates unstable derivatives [20]. For chiral analysis, by substituting a chiral thiol reagent for 2-mercaptoethanol, diasteromeric derivatives are produced and can be separated by HPLC [16].

Another technique that is routinely used to detect changes in the metabolic profiles is CE. Some reviews described advances in amino acid analysis by CE [21–23]; according to some of them, LIF provides the lowest reported LODs among the detection methods available for use with CE [24]. Several labeling techniques can be used with CE-LIF; each of these has its own strength and drawbacks. As reviewed previously [25], six different labeling agents are commonly used: NDA (naphthalene dicarboxaldehyde) [26], OPA [27], FQ (3-(2-furoyl)quinoline-2-carboxaldehyde) [28–30] and fluorescein derivatives such as FITC (fluorescein-5-isothiocyanate) [31, 32] or carboxyfluorescein succinimidyl ester [33]. The main drawbacks are the use of toxic chemicals such as sodium cyanide (NDA, FQ, 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA)), the expensive UV laser (OPA, NDA) [34], and the generation of a large number of fluorescein related labeling by-products. Moreover, FQ derivatives are not soluble enough in aqueous buffers [35]. Also, NDA, OPA, CBQCA, and FQ only react with primary amines; proline, a secondary amine cannot be labeled with these reagents.

4-Fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F) is a labeling agent that can be used for both primary and secondary amines. Sensitive detection of NBD-derivatives can be performed using HPLC and CE coupled to common argon ion laser [36]. When compared to fluorescein-based dyes, the advantages of using NBD-F are the short time required to achieve derivatization and the cleaner electropherograms due to low number of by-products. Tseng et al. [37] described a MEKC method based on NBD-F with a BGE containing sodium cholate, β-cyclodextrin (CD), Brij 35 in aqueous borate buffer, pH 9.3 containing 7% methanol; they applied this method to quantify amino acids in biofluids, and this includes a complex extraction procedure and complicated electropherograms. We recently validated a bioanalytical method based on CE-LIF using NBD-F as a derivatizing agent for the analysis of amino acids in urine and hippocampus tissue [11]. To our knowledge, there are no other published methods for the analysis of amino acids in plasma or serum by CE-LIF using NBD-F. We have also quantified the L-and D-serine enantiomers using the current methodology using the labeling agent, NBD-F. A previous study by Singh et al. [32] has used FITC to quantify selected amino acid enantiomers including serine. However, none of the previously reported CE based methods have been applied to study amino acid concentrations in plasma samples from psychiatric patients.

The goal of this work was to optimize and validate a method for analyzing amino acids in plasma by a CE method combining chiral selector and LIF detection. This validated method was applied to measure selected amino acids (L-proline, L-phenylalanine, L-leucine, L-isoleucine, L-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L-serine, D-serine, taurine and glutamate) using plasma samples of patients with BD.

2 Materials and methods

2.1 Chemicals and reagents

Analytical grade standards of L- and D-amino acids: phenylalanine 99%, glutamine 98%, serine 99%, aspartic acid 98%, glutamic acid 99%, valine, alanine, threonine, proline 99%, isoleucine 98%, leucine, ornithine hydrochloride 99%, 2-aminoadipic acid (internal standard—IS), glycine 99%, taurine 99%, β-CD 97%, and boric acid 99.5% were purchased from Sigma–Aldrich (Steinheim, Germany). Hydrochloric acid was from Fluka (Buchs, Switzerland) and sodium

hydroxide from Panreac Química (Barcelona, Spain). NBD-F 99% was from TCI (Tokyo, Japan). All solutions and dilutions were prepared with purified water from a Milli-Qplus 185 system (Millipore, Billerica, MA, USA). Individual 25 mM stock solution of each amino acid was prepared in purified water and stored at −20°C. From this, a dilution of 1 mM of each amino acid was prepared and stored at +4°C during the working week. These solutions were diluted as required on the day of the analysis. The derivatization solution of NBD-F was prepared by dissolving 80 mM NBD-F in methanol and mixing with equal parts of 500 μM HCl to give a final concentration of 40 mM NBD-F/250 μM HCl in 50% methanol.

2.2 Instrumentation

CE experiments were carried out on a P/ACE MDQ system (Beckman-Coulter, Fullerton, CA, USA) with LIF detector, an argon source operating at λexc: 488 nm and λem: 522 nm, a capillary silica column (Beckman Coulter, Madrid, Spain) 60 cm in total length, and 75 μm of id. All solutions were kept refrigerated at 7 ± 2°C in the CE autosampler. Data acquisition and instrument control were carried out using 32 Karat™ system software version 7.0 (Beckman Coulter, USA).

At the beginning of its use, the capillary was conditioned by flushing with 1 M NaOH (15 min) and water (15 min). Between the runs, the capillary was flushed with HCl 0.1 M (3 min), water (5 min), and BGE (5 min). Injections were made at the anodic end with a pressure of 0.5 psi (33 mbar) for 10 s. After optimization, running buffer (BGE) consisted of 175 mM borate buffer at pH 10.25 (pH adjusted with 2M NaOH) and 12.5mM β-CD. The voltage applied was +21 kV and the current observed under these conditions was 140 μA. The capillary thermostat was set at 17°C. The buffer vials were refreshed after every sixth analysis to maintain consistency.

2.3 Plasma samples

2.3.1 Study population

The study was approved by the Combined Neuroscience Institutional Review Board of the National Institutes of Health. All subjects provided written informed consent. All subjects were studied as inpatients at the National Institute of Mental Health Clinical Research Center, Mood Disorders Research Unit in Bethesda, Maryland. As described previously [38], patients corresponded to bipolar I or II disorder and were experiencing a major depressive episode without psychotic features were enrolled in the study. This study included nine subjects; one male and eight females, with an age range of 24 to 55 years (mean age = 39.2 (SD 10.5) years). These patients were also taking the mood stabilizer, lithium (0.6–1.2 mEq/L) for 4 wks before and during the study, no other psychotropic medications were taken. Pool of plasma for method development and validation was from healthy volunteers. All samples were stored at −80°C until the day of the assay.

2.3.2 Ultrafiltration

Plasma proteins were removed by filtration using the Centrifree® ultrafiltration devices from Millipore (Ireland; 30 kDa cutoff). A volume of 150 μL of plasma was placed into a single Centrifree® device and centrifuged at 2000 × *g*, 4°C for 60 min. The filtered plasma was then used for amino acid analysis.

2.3.3 Derivatization procedure

In a glass tube, 20 μL of filtered plasma, or the working standard solution was mixed with 20 μL of 200 mM aminoadipic acid (IS), 25 μL of 40 mM NBD-F (derivatization solution), and 150 μL of borate buffer 10 mM, pH 10. The resultant mixture was vortexed for 10 s and metabolite derivatization was performed at 60°C for 15 min in an oven. Immediately after derivatization, samples were kept refrigerated in the CE autosampler at 7°C for at least 30 min before CE-LIF analysis

2.3.4 Identification of amino acids in the plasma

Selected amino acids (L-proline, L-phenylalanine, L-leucine, L-isoleucine, L-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L-serine, D-serine, taurine and glutamate) were analyzed in this study. In order to identify and quantify these amino acids in plasma, the standard solution of the pure amino acid was injected alone, followed by the sample, and then the sample that was spiked with the standard. Moreover, L- and D- pure enantiomers for each amino acid were injected to check whether the final analytical method was capable of separating these isomers.

2.4 Validation study

The method was validated for selected amino acids for selectivity, linearity, accuracy, instrumental precision, method precision (both with standards and samples), LOQ, and LOD. Selectivity was checked by (i) analyzing the profile for ultrafiltered and nonderivatized plasma pool without the labeling agent; no peak was observed in this case; (ii) analyzing a blank containing the derivatizing reagent to distinguish the reagent peaks; and (iii) comparing the electropherograms of derivatized samples with and without the IS to check that there were no others peaks at IS migration time. Linearity was estimated by assaying at least five levels of concentrations of the standards in triplicate, covering all the expected values ranging from 25 to 200% or 300% of mean values found in a preliminary assay. The individual ranges are described in Table 1. Recovery was estimated by comparing in triplicate, the values of spiked samples prepared in a linear range (taking into account the endogenous concentrations that had been previously measured in the samples). Instrumental precision was evaluated by multiple injections ($n = 10$) of a homogeneous derivatized standard solution. Within-day precision of

Table 1. Validation data for selected amino acids in human plasma with the optimized CE-LIF method

	L-Proline	L-Isoleucine	L-Ornithine	L-Glutamine	L-Alanine	L-Threonine	Glycine	L-Serine	D-Serine	L-Glutamate
Linearity ($N = 20$)										
Slope	0.0078 ± 0.0003	0.0059 ± 0.0004	0.0017 ± 0.0002	0.0048 ± 0.0003	$0.00382 \pm 8.4 \times 10^{-5}$	0.0052 ± 0.0003	0.0059 ± 0.0004	0.0055 ± 0.0003	0.0047 ± 0.0002	0.00481 ± 0.00003
Intercept	-0.075 ± 0.099	-0.053 ± 0.067	-0.020 ± 0.028	-0.37 ± 0.27	-0.028 ± 0.038	-0.19 ± 0.12	-0.012 ± 0.018	-0.21 ± 0.13	0.0004 ± 0.003	0.012 ± 0.012
r	0.997	0.992	0.991	0.997	0.999	0.997	0.995	0.996	0.998	0.9997
Range (μM)	25–300	25–300	25–300	125–1500	25–750	25–600	25–750	25–750	0.5–25	25–300
Accuracy										
%	97.1	101.5	97.1	101.8	94.0	96.0	96.7	103.7	100.9	92.0
RSD(%)	3.7	5.2	6.9	4.5	7.8	7.3	6.0.	6.2	4.5	9.8
Instrumental precision ($N = 12$) RSD(%)	4.3	2.2	4.3	5.2	1.4	5.9	4.3	5.6	4.8	1.6
Method precision standard										
Interassay ($N = 12$) RSD (%)	6.5	4.2	6.8	6.2	3.8	5.5	5.2	6.7	4.4	2.8
Intra-assay ($N = 6$) RSD (%)	3.1	3.5	2.9	3.2	3.6	3.7	3.7	3.3	2.7	3.0
Method precision sample										
Interassay ($N = 12$) RSD (%)	6.0	5.3	6.0	5.7	5.8	5.5	4.4	5.3	6.8	3.4
Intra-assay ($N = 6$) RSD (%)	5.4	5.2	4.0	5.6	2.0	2.9	2.8	5.1	5.0	2.0
LOQ (EURACHEM) (nM)	738	630	1.44 μM	146	889	714	128	174	231	203
LOD (nM)	221	189	433	43.7	267	214	38.3	52.2	69.5	61.0

the method was checked by injecting individual preparations of standards and samples in the midrange of the calibration curve. Intermediate precision was tested in the same way, but on a different day, with freshly prepared buffers and reagents. LOQ for the selected amino acids were estimated using the Eurachem method [39] by injecting six replicates of each standard at least four levels of concentration ranging from 0.02–5.0 μM for D-serine and 0.1–20 μM for the rest of the amino acids. LOQ was established by representing RSD of the six replicates versus concentration and interpolating the concentration corresponding to 10%. LOD was calculated by means of the relation LOD: (3/10) × LOQ and checked experimentally.

3 Results and discussion

3.1 Method optimization

The principal goal of this study was to develop and validate a robust and rapid CE-LIF method for the determination of amino acid profile in human plasma. The method was based on our previous CE-LIF method for the analysis of amino acids in urine and hippocampus tissue samples [11], with some optimizations for this new biological matrix, human plasma. This validated method was used to measure selected amino acids in patients with BD.

In order to analyze amino acid concentrations in human plasma by CE-LIF, several optimizations were performed. Plasma proteins were first removed as amino acid profiling requires initial plasma deproteinization in order to avoid capillary clogging and changes due to absorptions in the capillary wall. These procedures involve the use of organic solvents or strong organic acids. The use of acid is also associated with some disadvantages; the labeling reaction (before derivatization with AQC, FITC, or NBD-F etc.) is performed in an alkaline medium by the addition of borate buffer. The acidified sample requires much more buffer for pH switching, thus increasing the ionic strength of the sample. The advantages to be gained from adopting the strategy of ultrafiltration are mainly not introducing any effect on the physicochemical status of the free amino acid fraction in the plasma and not introducing any substances that could affect derivatization. Because of the reasons cited above and based on the data available in the literature [14, 40], ultrafiltration was selected as the method for sample deproteinization. Plasma diluted (1:1) with pure water was compared with undiluted plasma sample. Better sensitivity without signal saturation was observed with the undiluted sample (data not shown). As an initial screen, previously described CE-LIF method [11] was applied to the filtered plasma after derivatization, but an unknown compound co-eluted with L-serine (which was not observed in our previous method). Lower concentrations of borate buffer were tested (90, 70, and 50 mM); decreasing the buffer concentration, analysis time was lowered but resolution equal to zero was achieved, nevertheless L- and D-serine were resolved. Higher concentrations of borate buffers were assayed: 110, 125, 150, and 175 mM after 30 s of sample injection (Fig. 1). Different sample injection times ranging from 10, 20, 30, 40, and 50 s were also tested, large differences in sensitivity were found. At more than 30 s of sample injection, peaks such as D-serine had higher signal and lower deviation, but a higher risk of peak co-migration was observed near the derivatizing peak. Thus, the 30 s injection time was applied for D-serine, and 10 s sample injection was applied for all other amino acids analyzed. With the use of 110 mM borate buffer (Fig. 1A), L-serine showed a single peak suggesting that the co-eluting compound migrated at the same time. As can be seen from Fig. 1, L-serine was fully resolved with higher concentrations of buffer (175 mM; Fig. 1D) and the peak shape was good. Moreover, both glutamine and alanine were fully resolved when compared to 125 and 150 mM concentration (Fig. 1B and C). Migration time, number of theoretical plates, and resolution were calculated for these two critical pair of peaks (Table 2); best results for resolution >1.5 were obtained with 175 mM BGE.

Addition of methanol or ACN to the BGE (up to 20% v/v was studied as the fluorescence of NBD derivatives is very sensitive to the hydrophobicity of the BGE, but higher migration time in the whole profile and wider peaks were obtained. The final conditions for the optimized CE-LIF method were: running buffer 175 mM borate buffer at pH 10.25, 12.5 mM β-CD, 10 s of sample injection (33 mbar), L-2-aminoadipic acid (IS using ultrafiltered, undiluted plasma samples.

The last step before validation was the identification of the selected amino acids in the plasma. As described in the method section, migration times and peak areas were compared between the ultrafiltered sample, pure standard and spiked sample (Fig. 2). Fourteen amino acids that were identified in the plasma profile include, L-proline, L-phenylalanine, L-leucine, L-isoleucine, L-ornithine, D-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L-serine, D-serine, taurine, and glutamate. Finally, for the identified amino acids pure L- and D-enantiomers were injected to ensure that the method could separate the enantiomers. The enantiomers of all these amino acids were satisfactorily separated. As a chiral selector was included in the BGE, chiral separation of the standards with resolution ≥1.5 was achieved (data not shown).

3.2 Validation

A complete validation was performed for only a representative group of amino acids including, L-proline, L-isoleucine, L-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L- and D- serine, and L-glutamate. A summary of the validation parameters for the selected amino acids are shown in Table 1.

During validation, standards fit the linear model ($r >$ 0.99) for all amino acids and no bias was found for most of them excluding, L-glutamine, L-threonine, and L-serine. However, in spite of the bias, no practical consequence was seen in the recovery. The recoveries ranged from 92.0 to 103.7%

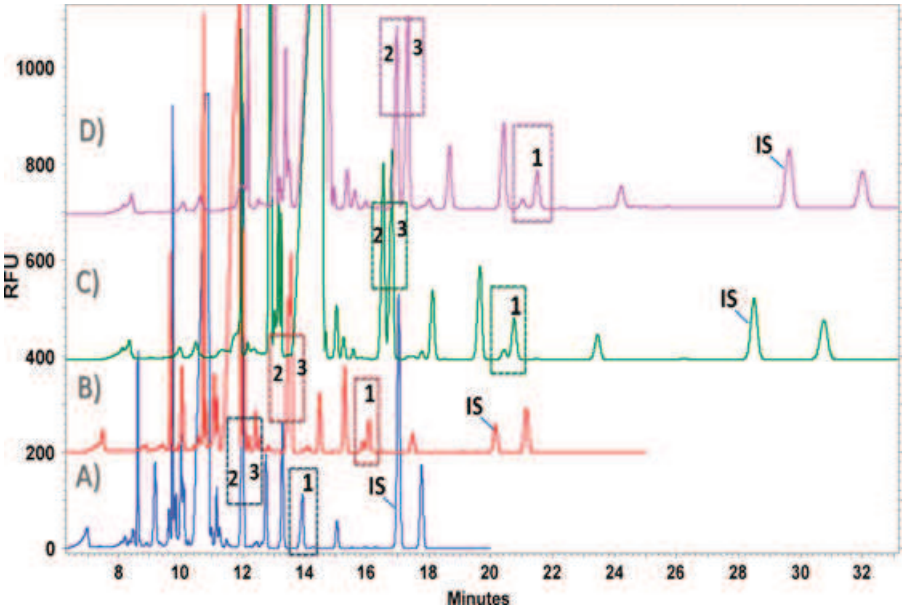


Figure 1. Influence of buffer concentration on amino acid profile; borate buffer, pH 10.25 and 12.5 mM β-CD, 30 s of sample injection. (A) 110 mM, [c.o. = 88 μA]; (B) 125 mM, [c.o. = 100 μA]; (C) 150 mM, [c.o. = 119 μA]; (D) 175 mM, [c.o. = 136 μA]. Key. [c.o.] = Current; [1] = L-serine; [2] = L-glutamine; [3] = L-alanine; [IS] = L-2-aminoadipic acid (A. 500 μM, B. 100 μM, C. and D. 250 μM).

Table 2. Performance parameters on critical separations of glutamine/alanine and unknown/L-serine through the increase of running buffer concentration

	Buffer 110 mM [c.o. = 88 μA]			Buffer 125 mM [c.o. = 100 μA]			Buffer 150 mM [c.o. = 119 μA]			Buffer 175 mM [c.o. = 136 μA]		
Analyte	tm (min)	*N(USP)*	*R(USP)*	tm (min)	*N(USP)*	*R(USP)*	tm (min)	*N(USP)*	*R(USP)*	tm (min)	*N(USP)*	*R(USP)*
Gln	12.0	139 113	0.0	13.5	29 080	0.3	15.5	160 923	1.1	17.0	101 646	1.7
Ala	12.0			13.6	104 822		15.7	161 776		17.3	98 086	
Unknown	13.9	110 615	0.0	15.9	148 677	1.1	18.9	140 663	1.1	21.0	94 944	1.7
L-Ser	13.9			16.1	155 873		19.1	151 890		21.5	95 162	

The amino acids with higher concentration range from 220–590 μM were L-glutamine, L-alanine, and glycine. Amino acids such as L-threonine, L-serine, L-ornithine, L-proline, L-isoleucine, taurine, and L-glutamate had a concentration range of 50 to 120 μM while D-serine had a concentration of approximately 2 μM. Abbreviations: Gln: glutamine; Ala: alanine; L-Ser: L-serine; tm: migration time; *N*: theoretical plates; *R*: resolution; [c.o.]: current observed.

and the differences were not statistically significant. For standards ($n = 12$), the instrumental precision ranged from 1.4 to 5.9%. Intra-assay precision for standards ranged from 2.9 to 3.7% ($n = 6$) and interassay precision from 2.8 to 6.8% ($n = 12$). Intra-assay precision for samples: six samples prepared from the same pool were treated from the beginning and run in the same assay, the daily RSDs ranged from 2.0 to 5.6% and from 3.4 to 6.8% in different days ($n = 12$). Theoretical LOQ calculated by the Eurachem method for these amino acids ranged from 128 nM for glycine to 1.44 μM for L-ornithine, which were lower than the values observed in the samples.

3.3 Quantitation of amino acids in plasma from BD patients

The validated CE-LIF method was applied to study plasma samples from nine patients diagnosed with BD. Eleven amino acids were determined in all the profiles: L-proline, L-isoleucine, L-ornithine, L-glutamine, L-alanine, L-threonine, glycine, L- and D-serine, taurine, and L-glutamate. All the results were in μM range and are summarized in Table 3. Interestingly, five patients whose samples were collected from 2007 to 2009 showed lower concentration of L-glutamine and higher concentration of glutamate than the more recent samples. A possible explanation for this is that serum hydrolases may increase the free glutamic acid by breaking down glutamine [41] during storage. Although these samples were stored at −80°C, the length of storage (3–5 years) could have contributed to this effect in both amino acids. Therefore, only the most recent samples were considered for statistical analysis of L-glutamate and L-glutamine. The results obtained and values from the literature are presented in Table 4. Major amino acids with concentrations from 220–590 μM were L-glutamine, L-alanine, and glycine, amino acids at middle concentrations from 50 to 120 μM were L-threonine, L-serine, L-ornithine, L-proline, L-isoleucine, taurine, and L-glutamate, and minor amino acids such as D-serine had a concentration of around 2 μM. Even though previous studies analyzed plasma samples not only from BD but also from other depressed

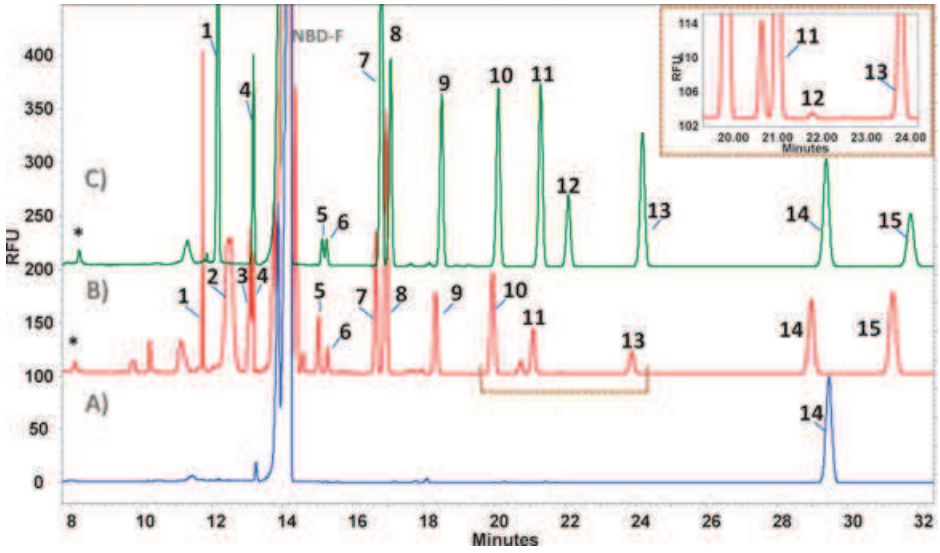


Figure 2. (A) Blank of sample (NBD-F). (B) Plasma profile. (C) Standards. Peak identification: 1. L-proline [200 μM]; 2. L-phenylalanine; 3. L-leucine; 4. L-isoleucine [100 μM]; 5. L-ornithine [100 μM]; 6. D-ornithine [100 μM]; 7. L-glutamine [500 μM]; 8. L-alanine [250 μM]; 9. L-threonine [200 μM]; 10. glycine [200 μM]; 11. L-serine [250 μM]; 12. D-serine [100 μM]; 13. taurine [150 μM]; 14. L-2-aminoadipic acid [IS = 200 μM]; 15. L-glutamate [100 μM]; D. NBD-F hydrolysis products. Conditions: 175 mM borate buffer, pH 10.25 and 12.5 mM β-CD, 10 s of sample injection.

Table 3. Quantitation of amino acids on plasma from BD patients

				Plasma concentrations (μM)				
Patient	Year	Gender	Age	L-Proline	L-Isoleucine	L-Ornithine	L-Glutamine	L-Alanine
1	2007	Female	34	70.9	60.0	121.8	476.8	349.2
2	2008	Female	24	36.1	38.7	60.5	403.0	268.1
3	2008	Male	55	67.1	63.3	94.6	560.8	399.9
4	2009	Female	37	73.1	75.4	78.9	471.0	440.1
5	2009	Female	44	63.7	68.0	113.5	354.9	322.5
6	2010	Female	48	41.8	46.2	77.4	633.7[a)]	298.3
7	2010	Female	29	23.0	43.3	54.8	531.8[a)]	278.8
8	2010	Female	32	56.9	45.4	75.0	433.7[a)]	214.4
9	2011	Female	50	57.4	42.7	81.7	698.2[a)]	261.1
			Average	54.5	53.7	84.3	574.3	314.7
			SD	16.2	12.4	21.0	116.1	67.4

		Plasma concentrations (μM)					
Patient	Sample year	L-Threonine	Glycine	L-Serine	D-Serine	L-Glutamate	Taurine
1	2007	136.7	209.3	125.2	1.3	130.5	57.0
2	2008	93.4	197.1	93.6	1.7	123.1	33.3
3	2008	134.7	251.7	126.4	2.3	85.5	124.8
4	2009	116.0	192.7	85.1	4.5	211.5	55.3
5	2009	99.9	209.3	92.5	1.9	193.1	79.8
6	2010	177.5	247.5	109.4	3.5	22.4[a)]	26.5
7	2010	120.2	380.1	94.3	2.7	27.5[a)]	55.6
8	2010	131.7	145.2	73.4	1.2	27.7[a)]	21.3
9	2011	135.4	254.5	110.0	2.8	35.5[a)]	24.0
	Average	127.3	231.9	101.1	2.4	28.3	53.1
	SD	23.2	61.8	16.9	1.0	5.4	31.3

a) Values of L-glutamine and L-glutamate considered for statistical analysis.

patients undergoing lithium or other treatments, our results are concordant with the literature. The glutamate and glycine results were very similar to the previous values reported for patients suffering from major depressive disorder [42] and BD patients [43]. Taurine values were lower compared with previously studied depressed patients [3, 44], but comparable levels were observed with BD patients [43]. For this instance antidepressant treatment may significantly decrease plasma taurine levels. Amino acid values obtained for L-alanine, L-glutamine, L-ornithine, L-isoleucine, L-threonine,

Table 4. Amino acids quantitation in plasma from BD patients ($n = 9$) and comparison with similar studies [3, 4, 42–44]

Amino acid (μmol/L)	Current results ($n = 9$)		Mitani et al. [3] ($n = 23$)		Hoekstra et al. [4] ($n = 5$)		Hoekstra et al. [4] ($n = 20$)		Mayoral-Mariles et al. [42] ($n = 17$–21)		Altamura et al. [43] ($n = 25$)		Pinto et al. [44] ($n = 5$)	
	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM	Average	SEM
L-Glutamine	574.3[a)]	50.3	393.4	40.4					583.4	28.4	447.0	65.2		
L-Alanine	314.7	23.8	404.4	42.4									273.0	11.0
Glycine	231.9	21.9	205.0	23.5	319.2	47.3	288.8	20.9	240.4	13.5	215.0	10.4	159.0	11.0
L-Threonine	127.3	8.2	155.2	17.2					80.9	5.4			84.0	2.0
L-Serine	101.1	6.0	107.4	9.6					93.1	3.6	108.0[b)]	4.2	153.0[b)]	9.0
L-Ornithine	84.3	7.4											97.0	2.0
L-Proline	54.5	5.7												
L-Isoleucine	53.7	4.4							44.9	2.5			71.0	11.0
Taurine	53.1	11.1	133.9	19.4							40.0	2.6	147.0	6.0
L-Glutamic acid	28.3[a)]	2.7	94.0	10.0	39.4	8.6	56.8	6.1	44.1	3.3	336.0	31.8	58.0	2.0
D-Serine	2.4	0.4	2.1	0.2										
Subject details	Inpatients with a diagnosis of bipolar I or II depression under lithium treatment		Patients with depression (17 out of 23 treated with imipramine drug)		Manic patients with bipolar I disorder under lithium		Manic patients with bipolar I disorder		Mild depressed elderly women		Bipolar and unipolar major depressed subjects		Patients with major depression diagnosed of anxiety disorder	

a) L-glutamine and L-glutamate values from 4 patients (collected between 2010–2011). Data presented are mean ± SEM.
b) Quantification of total serine.

and L-serine, reported here are in agreement with the literature [3, 4, 42–44]. Interestingly, most previously published articles reported total serine instead of the separate enantiomers.

4 Concluding remarks

We developed a sensitive CE-LIF method for the analysis of 14 amino acids in human plasma using 175 mM borate running buffer at pH 10.25, 12.5 mM β-CD, 10 s of sample injection (33 mbar) and L-2-aminoadipic acid as the internal standard. The method has been validated for 10 amino acids including some D- and L-enantiomers such as D- and L-serine, which cannot be detected with other general methods based on GC and HPLC. Validation parameters are adequate for bioanalysis. Undiluted human plasma samples can be processed by ultrafiltration, derivatized, and analyzed for amino acid profiles. The results obtained for representative amino acids plasma from BD patients are in agreement with literature values. The method is useful particularly in studies where plasma amino acid levels in patients are used as biomarkers for diagnosis of diseases, evaluating the disease progression, and monitoring response to drug therapy.

The authors thank Irving W. Wainer and Carlos A. Zarate, for the plasma samples and for their irreplaceable support and also Santiago Angulo for his scientific support. Alma Villaseñor acknowledges EADS-CASA for her fellowship. The authors gratefully acknowledge the financial support from Ministry of Science and Innovation MICINN CTQ2011-23562. This work was supported in part by the Intramural Research Programs of the National Institute of Aging, National Institutes of Health (NIH), and the National Institute of Mental Health, NIH.

The authors have declared no conflict of interest.

5 References

[1] He, Y., Yu, Z., Giegling, I., Xie, L., Hartmann, A. M., Prehn, C., Adamski, J., Kahn, R., Li, Y., Illig, T., Wang-Sattler, R., Rujescu, D., *Transl. Psychiatry* 2012, *2*, e149.

[2] Xu, H. B., Fang, L., Hu, Z. C., Chen, Y. C., Chen, J. J., Li, F. F., Lu, J., Mu, J., Xie, P., *Psychiatry Res.* 2012, *200*, 1054–1057.

[3] Mitani, H., Shirayama, Y., Yamada, T., Maeda, K., Ashby, C. R., Kawahara, R., *Prog. Neuropsychopharmacol. Biol. Psychiatry* 2006, *30*, 1155–1158.

[4] Hoekstra, R., Fekkes, D., Loonen, A. J., Pepplinkhuizen, L., Tuinier, S., Verhoeven, W. M., *Eur. Neuropsychopharmacol.* 2006, *16*, 71–77.

[5] Zarate, C. A., Manji, H. K., *Mt. Sinai J. Med.* 2008, *75*, 226–247.

[6] Sussulini, A., Prando, A., Maretto, D. A., Poppi, R. J., Tasic, L., Banzato, C. E., Arruda, M. A., *Anal. Chem.* 2009, *81*, 9755–9763.

[7] Waziri, R., Wilson, R., Sherman, A. D., *Br. J. Psychiatry* 1983, *143*, 69–73.

[8] Waziri, R., Wilcox, J., Sherman, A. D., Mott, J., *Psychiatry Res.* 1984, *12*, 121–136.

[9] Maes, M., De Backer, G., Suy, E., Minner, B., *Neuropsychobiology* 1995, *31*, 10–15.

[10] Zinellu, A., Sotgia, S., Pisanu, E., Scanu, B., Sanna, M., Usai, M. F., Chessa, R., Deiana, L., Carru, C., *Anal. Bioanal. Chem.* 2010, *398*, 1973–1978.

[11] Lorenzo, M. P., Navarrete, A., Balderas, C., Garcia, A., *J. Pharm. Biomed. Anal.* 2013, *73*, 116–124.

[12] Martens-Lobenhoffer, J., Bode-Böger, S. M., *Clin. Chem.* 2006, *52*, 488–493.

[13] Sethuraman, R., Krishnamoorthy, M. G., Lee, T. L., Liu, E. H., Chiang, S., Nishimura, W., Sakai, M., Minami, T., Tachibana, S., *Clin. Chem.* 2007, *53*, 1489–1494.

[14] Jaworska, M., Stańczyk, M., Wilk, M., Kłaczkow, G., Anuszewska, E., Barzał, J., Rzepecki, P., *Amino Acids* 2012, *43*, 1653–1661.

[15] Frank, M. P., Powers, R. W., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2007, *852*, 646–649.

[16] Grant, S. L., Shulman, Y., Tibbo, P., Hampson, D. R., Baker, G. B., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2006, *844*, 278–282.

[17] Shimbo, K., Oonuki, T., Yahashi, A., Hirayama, K., Miyano, H., *Rapid Commun. Mass Spectrom.* 2009, *23*, 1483–1492.

[18] Kaspar, H., Dettmer, K., Chan, Q., Daniels, S., Nimkar, S., Daviglus, M. L., Stamler, J., Elliott, P., Oefner, P. J., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2009, *877*, 1838–1846.

[19] Fekkes, D., *J. Chromatogr. B Biomed. Appl.* 1996, *682*, 3–22.

[20] Devall, A. J., Blake, R., Langman, N., Smith, C. G., Richards, D. A., Whitehead, K. J., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2007, *848*, 323–328.

[21] Poinsot, V., Gavard, P., Feurer, B., Couderc, F., *Electrophoresis* 2010, *31*, 105–121.

[22] Poinsot, V., Rodat, A., Gavard, P., Feurer, B., Couderc, F., *Electrophoresis* 2008, *29*, 207–223.

[23] Iadarola, P., Ferrari, F., Fumagalli, M., Viglio, S., *Electrophoresis* 2008, *29*, 224–236.

[24] Lapainis, T., Sweedler, J. V., *J. Chromatogr. A* 2008, *1184*, 144–158.

[25] Szöko, E., Tábi, T., *J. Pharm. Biomed. Anal.* 2010, *53*, 1180–1192.

[26] Siri, N., Lacroix, M., Garrigues, J. C., Poinsot, V., Couderc, F., *Electrophoresis* 2006, *27*, 4446–4455.

[27] Taga, A., Honda, S., *J. Chromatogr. A* 1996, *742*, 243–250.

[28] Veledo, M. T., de Frutos, M., Diez-Masa, J. C., *J. Chromatogr. A* 2005, *1079*, 335–343.

[29] Boulat, O., McLaren, D. G., Arriaga, E. A., Chen, D. D., *J. Chromatogr. B Biomed. Sci. Appl.* 2001, *754*, 217–228.

[30] Bergquist, J., Vona, M. J., Stiller, C. O., O'Connor, W. T., Falkenberg, T., Ekman, R., *J. Neurosci. Methods* 1996, *65*, 33–42.

[31] Nouadje, G., Rubie, H., Chatelut, E., Canal, P., Nertz, M., Puig, P., Couderc, F., *J. Chromatogr. A* 1995, *717*, 293–298.

[32] Singh, N. S., Paul, R. K., Sichler, M., Moaddel, R., Bernier, M., Wainer, I. W., *Anal. Biochem.* 2012, *421*, 460–466.

[33] Chen, H. L., Zhang, X. J., Qi, S. D., Xu, H. X., Sung, J. J., Bian, Z. X., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 2009, *877*, 3248–3252.

[34] Zhang, H., Le Potier, I., Smadja, C., Zhang, J., Taverna, M., *Anal. Bioanal. Chem.* 2006, *386*, 1387–1394.

[35] Klinker, C. C., Bowser, M. T., *Anal. Chem.* 2007, *79*, 8747–8754.

[36] Hu, S., Li, P. C., *J. Chromatogr. A* 2000, *876*, 183–191.

[37] Tseng, H. M., Li, Y., Barrett, D. A., *Anal. Bioanal. Chem.* 2007, *388*, 433–439.

[38] Diazgranados, N., Ibrahim, L., Brutsche, N. E., Newberg, A., Kronstein, P., Khalife, S., Kammerer, W. A., Quezado, Z., Luckenbaugh, D. A., Salvadore, G., Machado-Vieira, R., Manji, H. K., Zarate, C. A., *Arch. Gen. Psychiatry* 2010, *67*, 793–802.

[39] Eurachem Working Group, *The Fitness for Purpose of Analytical Methods*, 1998, http://www.eurachem.org//images/stories/guides/pdf/valid.pdf

[40] Jensen, B. P., Chin, P. K., Begg, E. J., *Anal. Bioanal. Chem.* 2011, *401*, 2187–2193.

[41] Deyl, Z., Hyanek, J., Horakova, M., *J. Chromatogr.* 1986, *379*, 177–250.

[42] Mayoral-Mariles, A., Cruz-Revilla, C., Vega-Manriquez, X., Aguirre-Hernández, R., Severiano-Pérez, P., Aburto-Arciniega, E., Jiménez-Mendoza, A., Guevara-Guzmán, R., *Arch. Med. Res.* 2012, *43*, 375–382.

[43] Altamura, C., Maes, M., Dai, J., Meltzer, H. Y., *Eur. Neuropsychopharmacol.* 1995, *5*, 71–75.

[44] Pinto, V. L., de Souza, P. F., Brunini, T. M., Oliveira, M. B., Moss, M. B., Siqueira, M. A., Ferraz, M. R., Mendes-Ribeiro, A. C., *J. Affect. Disord.* 2012, *140*, 187–192.

Chapter 7

7. Discussion

The field of metabolomics was born as the key to connect the gap between metabolite changes and observed phenotypes. In a way, this new field was the evolution of classical biochemistry, where the measurement of specific compounds was performed with the purpose of covering different clinical goals. Among the goals, the quantification of a generally small set of compounds would support a diagnostic or prognosis tool, or the establishment a clinical frame, but these purposes were mostly unsatisfactory or failed in fulfilling such tasks. At that time the analytical techniques available were capable of detecting only very few biological compounds, and only in medium to high concentrations in the samples. As very often happens, instrumental development brought a change in the paradigm. This initiated a big success in the biological field, where the biological hypotheses were beginning to be challenged, and even biological "*facts*" were brought down.

Concerned about the importance of the metabolites and the advances of technologies, researchers looked at the analytical techniques and established the basis of the new emerging discipline of metabolomics. Here, analytical techniques became one of the mainstays in the field. For that reason, getting good knowledge about strengths and drawbacks for approaching different problems was the focus of this dissertation, and most of the analytical techniques currently used in metabolomics were covered.

As it was said from previous chapters, there is no single technique or methodology capable of detecting and quantifying all metabolites in a biological sample. Therefore, to catch and look at different parts of the metabolome, diverse analytical platforms and strategies have been developed. In addition, the idea of using a combination of different analytical techniques to increase the metabolome coverage has got certain fame, but the complete integration of the results still is not totally clear. Of note to highlight is that each analytical technique is defined by its capabilities, with strengths and weak sides. Conscious of this, their application in the metabolomics field has to be done by taking the best of each, and choosing the best way possible way around each drawback.

GC–Q–MS was the first MS–based technique used in metabolomics and was, at the beginning, used a lot on plant extracts and as it was suitable for detecting volatile and thermally stable compounds. Over time, samples such as urine and plasma were made possible to be measured in GC by the improvement of derivatization methods. Finally, at the present moment, the GC–Q–MS process has become a reproducible technique that can allow the robust acquisition of non–target profiles in biofliuds. Particularly, the characterization of the metabolome from biofluids such as urine and plasma is now well extended. Nevertheless, the characterization of other biological matrices is an important task in other biofluids such as breast milk (BM, Chapter 2). Here the development of analytical methods and their validation is the first step forward.

The analysis of the components of BM is not easy: the majority of compounds are lipids and the reference extraction methods Folch, and Bligh and Dyer depicted involve hard and extensive manual work. Therefore, the need to look for new extraction protocols seems necessary. New extraction system employing methyl *ter*–butyl ether (MTBE) and methanol (1:1) showed the huge advantage of obtaining one phase instead of using a typical two phase system. This optimization made it possible to: considerably reduce sample preparation (to a few minutes per system); avoid the use of chloroform and the partition mechanism of the molecules when two phases; and get a high recovery. The optimization was tested by LC–QTOF–MS and once the protocol was established, the extracted sample was characterized by LC–MS and GC–MS. Within these, GC–MS was used with the simplest mass analyser: the –Q. Here it is important to highlight that the extraction protocol will define what metabolites will be present in the extracted sample, while the type of metabolites and their detection is delimited by the analytical platform.

Among the limitations observed during the BM characterization, we found that the filament from the EI source got dirty quickly, and thus decreased the instrument's sensitivity. This process was caused by the presence of high concentrations of disaccharides in the extracted BM system. Therefore, the method of analysis in the instrument was modified, turning the filament off during this band of time (within the minute 23 to 26). In addition the use of Q–analysers for the identification of metabolites which don't have their fragmentation patterns included in the spectra databases is a hard task. This is because the Q is a low resolution mass analyser and so the interpretation of the fragments is highly ambiguous. An attempt to resolve this was made by the coupling of QTOF mass analysers to GC, as TOF instruments provide high resolution and mass accuracy allowing the opportunity to assign identification.

For this application, GC–Q–MS for BM showed the identification of 60 metabolites in the total profile, including amino acids, organic acids, free fatty acids, hexose and pentose sugars, TCA intermediates, cholesterol and disaccharides. Although, it was limited to only detect metabolites that are or can become volatile and thermal stable. The use of EI as an ion source in this type of instrument showed an advantage of providing a highly confident identification of metabolites. This is because EI has reproducible fragmentation patterns with a minimal variation among instruments, allowing for the possibility of identifying metabolites, not only by fragmentation pattern, but also by retention index (RI) score when a standard methodology is applied. So while in other analytical platforms the identification is one of the last steps in the workflow (such as in LC–MS), GC–MS uses the identification process as one of the first steps in the data processing.

There is not an established guideline on how to validate non–targeted methods, so among the possibilities discussed in Chapter 2, the validation of traditional parameters (selectivity, calibration model (linearity and range), accuracy, precision (repeatability and intermediate precision), and limit of detection (LOD) and quantification (LOQ)) was chosen. Here, 18 compounds were selected between LC–MS and GC–MS platforms, and from these the stearic acid was common for both. This showed the same concentration after its quantification in the pool of BM samples, highlighting a predictor of the extraction and technique robustness.

The use of analytical techniques in metabolomics started with technological advances in NMR – a known technique for providing structural information (a fundamental aspect in metabolomics). This was the first analytical platform that caught the eye of researchers and commercial companies. The structural elucidation of NMR in combination with the increment of magnetic strength from improved technology, quickly made NMR a high throughput technique. With the increase in sensitivity obtaining "*metabolic non–target profiles*" from biofluids such as urine was successful. The urinary 1D spectrum was an important accomplishment in metabolomics, where hundreds of metabolites were obtained in a fast and single run.

NMR was for a long time the workhorse in the metabolomics field. Here its characteristics as a non–destructible and highly reproducible technique capable of detecting specific resonance adsorption, typically from ^{1}H nuclei in metabolites in the concentration range of μM, were highly exploited in many

research areas. The capability of measuring biological samples with minimum or no pretreatment was also very important, and very close to the ideal of metabolomics. This was especially straightforward with urine samples because, unlike plasma, the low content of proteins permitted the extraction of a clear spectrum with thousands of signals. For plasma, however, the content of proteins, lipids and macromolecules caused the presence of broad resonances in the spectrum which, along with the high concentration of some molecules (e.g. glucose), gave a high overlap for low weight metabolites. Even when this was later overcome by using the spin–echo experiment CPMG to attenuate signals from broad macromolecular components enabling a much clearer profile, the rich information in the spectrum observed with urine was still not reached. Importantly, the creation and development of bioinformatics and statistical spectroscopy tools and approaches for complex NMR data treatments were decisive in establishing and strengthening the data workflow, and getting reliable and valuable results for data interpretation. Worth mentioning is the STOCSY tool that allows the correlation of one signal in the profile across all signals in the spectrum, giving chemical shifts corresponding to the same molecule and, as a result, highly confident identification.

Among the challenges and troubles faced by NMR, the high impact of the presence of xenobiotics (e.g. drugs or excipients) on the data, or the high content of protein owing to disease can be found in urine samples, and therefore can cause bias or difficulties in the analysis (Chapter 3). This fact caused urine samples with a high protein content to be removed from the analysis and the study, bringing a decrement in the force of the statistical analysis due to lower sample number in the comparison.

In the case of the presence of xenobiotics in the NMR profile, these in some cases do not impact the data analysis i.e. they are not statistically significant for the analysis; even if they show several or high signals in the profile, their elimination is not compulsory. But when the xenobiotic compound becomes discriminant in the analysis the only way to eliminate their influence from the study is by cutting the zones from the profile that span the chemical shifts of these xenobiotic signals in the raw data. Here, the presence of xenobitiocs could mean cutting a complete segment of the profile if the xenobiotic span signals in the profile overlap other signals (e.g. mannitol overlapping almost all the zone of sugars), or shows several peaks (e.g. acetaminophen and its related metabolites). Although means that the compounds that are overlapped or close to xenobiotic signals get lost from the analysis, this then permits the collection of real biomarkers for the comparison. Importantly, even with these problems with NMR,

this technique is still one of the principal platforms used, based on its advantages and capabilities which contribute to the development of metabolomics.

Along with the NMR platform, classical analytical techniques were explored with the possibility of being included in the list of platforms in the metabolomics arena. Among them, CE was a very promising technique which seems to be more than suitable for urine analysis. CE was characterized to be one of the highest resolution techniques that allowed the analysis of an almost intact urine sample (only diluted with water). Of course, its main drawbacks are enhanced when UV detection is used, because it does not provide structural information, has poor sensitivity (due to the small volume sample and short optical path–length of the capillary), is not completely universal, and has poor migration time reproducibility from CE itself. Nevertheless, this technique has shown its value based on the advantages that it represents. The development and validation of two previously methodologies, one with reverse polarity on CZE mode for ions and the other in CD–MEKC mode for cations and neutrals, could allow the analysis of urine in both modes with only 15µL of urine sample (Chapter 4). The control and the optimization of parameters of CE during the method development allowed a good reproducibility of the migration time that could mostly be corrected by the creation of alignment algorithms such as the COW algorithm.

The data combination of NMR and CE–UV (both complementary techniques with sharing metabolites) aroused the possibility to not only extends metabolite coverage, but also the opportunity of identifying a CE–UV signal through the high correlation shown with another signal in the NMR spectrum. The extension of the metabolite coverage by combining the profiles from the two techniques encouraged the study of correlations by taking one signal in the profile and getting connections with other signals from the now bigger profile. In this way, more information on specific molecules was obtained where possible pathway associations with other metabolites were established (e.g. for oxalate in the slc26a6 knockout model, Chapter 4). For metabolomics, the purpose of identification is necessary to being able to explain the metabolic changes further, and to get a better understanding of any project case. Therefore, this made researchers stop using CE–UV for metabolomics. Nevertheless, personally, above all the great characteristics and advantages of this technique, the small cost that this represents should make CE–UV an actual technique used in the third world countries, where the possibility to reach expensive apparatus is not affordable. One of the options here is the possibility of transferring the results

(biomarkers) of specific diseases by previous work by joining NMR and CE–UV data. In addition, it is important to highlight the low or non–existent use of organic solvents in the CE–UV technique.

Nevertheless, the interest in uncovering the metabolites under the sensitivity range of NMR (and also CE–UV) made the exploration of new analytical routes necessary. Among other difficulties faced with NMR such as the complexity data handling, overlapping of compounds or simply because the compound could not be detected by this technique, the research group was moved to use MS platforms.

New analytical platforms introduced were then conditioned by the ability of the technique to get identification of the metabolites of interest, and thus MS was enlisted and adopted in the field. The direct infusion in MS caused the surprising collection of such a huge amount of information at high sensitivity. But the handling of such complex data, along with technical problems such as ion suppression and the compromise in mass resolution, created the possibility of coupling MS to separation techniques, where GC, LC and finally CE were used.

The improvements with each MS–hyphenated technique were dependent on technology advances by the instrument manufacturers in aspects such as the interfaces, mass analysers robustness, and the development of vacuum turbo–molecular pumps. Since then, LC–MS was probably the technique most improved by the commercial brands because of its high capabilities and contributions to metabolomics. Here, different instrumental modalities have been tested and studied, such as various ion sources: API (i.e. ESI) which is the most used source in metabolomics studies; and for other applications, APPI, APCI and MALDI; mass analysers IT, Q–TOF, FT–ICR, QqQ and detectors. Here while positive ESI ionization is capable of detecting the widest range of metabolites, negative ionization mode is commonly used to extend the metabolite coverage. Therefore their combinations and the type of mass analyser grant the characteristics and capabilities for each LC–MS instrument. In addition, LC–MS has received most of the researchers' attention, who have tested and developed an extensive number of methodologies and approaches, such as distinct modes of chromatography (e.g. RP and HILIC), chromatographic gradients, and types of buffers, etc. Along with all these possibilities, big efforts were made to improve MS data analysis through new software and programs for peak picking, mass alignment, identification of compounds, pathway analysis, and multivariate data treatment.

LC–MS is an instrument that requires careful handling, checking allinstrumental parameters constantly, e.g. the curve of pressure, signal intensity, and reference masses signal. In addition to monitoring the performance of data acquisition, the hyphenated MS–based techniques need to use QC samples in order to prove the reliability of data analysis and the results obtained. So the use of QC samples is necessary for any non–targeted analysis sequence, because MS instruments have an extremely high sensitivity, and the appearance of ghost peaks and high noise is highly probable. The quality of performance can be established when the QC samples (after filtering) get clustered into a no supervised model, such as a PCA plot. This guarantees to keep all the features from QC samples, where signal intensity is stable during the complete analysis and the masses considered valid have typical RSDs of 20–35%. As said before, the characteristics of LC–MS analysis are based on all of the instrument and method conditions. For the project in Chapter 5, the LC–MS instrument employed had an ESI ion source and a Q–TOF mass analyser. This was used only in positive mode, and the run method considered RP chromatography using a C18 column. This meant that the analytes detected by this method, although determined by the extraction protocol of the sample, covered a broad range of molecular weights (the MS parameters set in the instrument method which are typically from *m/z* 100 to 1 000Da) and separated from medium to non–polar compounds. Here, the first characteristic of the LC–MS is that the sample needs to be treated (or extracted) and based on the protocol followed, the metabolite yield is different. Nevertheless I consider that part of the problems associated with LC–MS are currently in the data treatment context. Here, the data manipulation approaches have not been completly worked out, e.g. how to filter the data avoiding bias, how to treat missing values, logarithmic transformation and the establishment of a normalization method for MS–based data. In addition, the final identification of significant metabolites is lessened, because of the lower number of compounds that can be finally confirmed with a reagent standard (Figure 5.6, Chapter 5). The identification process faces several challenges, and these are displayed in the next list:

1. The databases (METLIN, KEGG, HMDB, etc.) do not contain all metabolites, and from these only a fraction contains the fragmentation spectra. Moreover, in order to have a reliable comparison with the fragmentation spectra from database (METLIN), the MS/MS spectrum has to be acquired with the same type of MS instrument, and if possible, identical or close conditions.

2. The molecular fragmentation in the ion source is quite frequent, and this results in misleading interpretations.
3. The MS/MS experiments can fail because of several reasons: (*i*) if the molecule cannot be fragmented because it has a hard structure or is clustered to a metal ion like sodium; (*ii*) if there are two masses with very close *m/z* values co–eluting at a very close RT, it is highly probable that the MS/MS spectrum will not be clean; (*iii*) similar to the previous, high noise signals will overlap the fragment signals, and thus their isotope distribution; (*iv*) a high amount of fragments makes their analysis and interpretation difficult.
4. MS/MS fragments can be misinterpreted or simply not fully explained.
5. Finally, not all the standards are commercially available.

Nevertheless the identification process by MS is improving, and new strategies and research in the fragmentation process are providing advances. Despite the drawbacks from LC–MS, this has valuable advantages in research, based on its high sensitivity and broad range of biochemical classes of metabolites that is possible to separate and detect. This has made LC–MS to be one of the most important techniques in metabolomics.

For non–targeted analysis of biofluids, techniques such as NMR, CE–UV and LC–MS (Chapter 3 to 5) show different advantages, capabilities and limits according to the method of analysis and metabolites detected, and the scope of each analytical platform. In addition, other analytical platforms that are currently used for the same purpose are GC–MS and CE–MS. Meanwhile, for target analysis, the use of analytical platforms that could allow high selectivity for the metabolites of interest is highly desirable. In addition, other important needs are high analysis speed, sensitivity and resolving power. In this sense, the use of QqQ in the mode of multiple reaction monitoring (MRM) and selected reaction monitoring (SRM) are the most highly technological and frequently used methodologies for target analysis. Nevertheless, when amino acids are the target, the CE–LIF technique is a good choice (Chapter 6). Thus CE–LIF is characterized by the combination of a high resolution separation achieved by CE and LIF as one of the most sensitive detectors and highly selective because of its potential to detect native fluorescence or by labelling the compounds of interest with a fluorophore.

Within the different type of dyes and laser systems developed for plasma treatment, the labelling reagent NBD–F and the ultra–filtration devices for sample protein removal were chosen. NBD–F reagent has shown its advantages for primary and secondary amines and its sensitive detection by argon ion laser. The combination of both aspects allowed for the development of an easy methodology (that takes only 15 minutes for the derivatization reaction) and that is able to detect 14 amino acids in plasma, where all (L– and D–) enantiomers were satisfactorily separated. The method was completely validated and it was applied to the characterization of amino acids in treatment-resistant patients with bipolar disorder. Here again one of the big advantages of using CE–LIF compared to tandem MS such as QqQ for target analysis, is based on the low cost of the technique and that the results (at least for amino acids) do not seem compromised.

In summary, each analytical platform presents different characteristics which make them suitable for a specific set of metabolites. Figure 7.1 was created to illustrate that the capabilities of analytical techniques to characterize the metabolome. In this context, the metabolism of any organism (represented by human in the figure) is like a puzzle where each technique depends on the physicochemical properties of the metabolites and the limit of concentration range (NMR, CE–UV < MS) will catch different pieces (with some in common). Nevertheless, even using MS–based platforms, there will still be compounds that are below MS detection limit (represented by the feet of the man in the picture). Here NMR, CE–UV, LC–MS, and GC–MS can work as reliable tools to obtain the non–target profile from biological samples. This picture also shows that GC–MS is one of the techniques that can be used to characterize the non–target profile of a new biological matrix (represented here by the man's hair). Moreover, when a metabolite is found altered by a physiological condition and target analysis is needed, techniques such as CE–LIF can be of great advantage based on its specificity and high sensitivity. All these analytical platforms become part of the tools to obtain an integrated profile (metabolome).

Finally, the metabolomics field is in continuous evolution and along with it, the analytical platforms and informatics tools. With this in mind, is possible that near new diagnostic and prognostic tests are going to be made by any of the high throughput techniques.

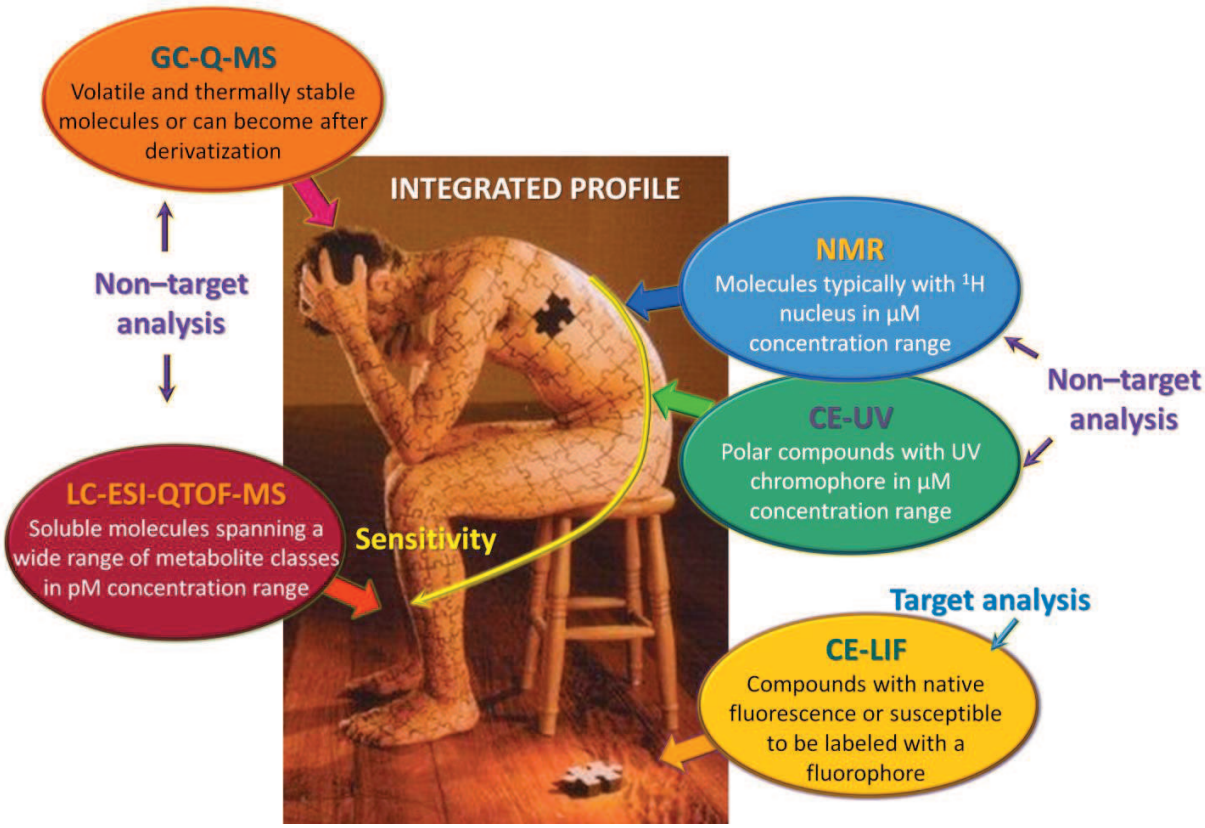


Figure 7.1. Analytical techniques and metabolite characteristics in each platform for metabolic non–target (NMR, CE–UV, LC–ESI–QTOF–MS and GC–Q–MS) and target analysis (CE–LIF). Puzzle picture taken from http://en.wikipedia.org/wiki/Puzzle_(Biffy_Clyro_album).

CONCLUSIONS

This work has highlighted the importance of analytical platforms in metabolomics. Although there is currently no single technique that is able to measure all metabolites in an organism, the way that each analytical technique contributes to solve biological questions according to their capabilities has been shown. Here, not only a complete description of the characteristics (advantages and limitations) of the most relevant analytical techniques employed in metabolomics was performed, but also the diverse concerns and challenges related to them were discussed. For each of them, their use, and the subsequent treatment of data generated by them to reach a biological interpretation, was described.

The paramount importance of the analytical techniques GC–MS, NMR, CE–UV, LC–MS and CE–LIF was proved over the different biological applications within each strategy in metabolomics: method development and validation; metabolic non–targeted; and targeted analysis. In addition, through these chapters, the evolution of analytical techniques used for metabolomics can be followed, from initial equipment (e.g. CE–UV) to second generation (e.g. NMR and GC–MS) or newer instrumentation (e.g. LC–MS). Moreover, the combination or complementarity among different techniques has been shown.

Particular conclusions of this work are listed by chapters:

1. An in–depth study of the literature for sample non–targeted analysis with CE either with UV or MS detection allowed a critical assessment of the valuable characteristics of CE for the analysis of different biological fluids, especially for urine samples due to the abundance of polar and ionic compounds. Among the strengths, high efficiency and resolution, small sample volume and minor sample treatment were the most significant, while on the other hand one of the main drawbacks was its lack of robustness (Chapter 1).
2. For the first time, the breast milk metabolome has been obtained based on one single phase extraction system, and characterized by two instrumental techniques: the GC–Q–MS with a high potential for the study of metabolites in biological samples that can be volatile or thermally stable after derivatization; and the LC–TOF–MS where different and complementary classes of

biochemical compounds were observed (Chapter 2). The methodologies on both instruments were fully validated, presenting acceptable values for biological studies, demonstrating its value for application in further studies. Moreover, evolution of breast milk over time was obtained, confirming previously known changes. The fatty acids and cholesterol were significant in both techniques following the same pattern, while glycerolipids and phospholipids were from LC-MS, and sugars and carbohydrates from GC-MS.

3. The use of non-targeted analysis by NMR deployed a data treatment–strategy to determine a diagnostic and prognostic phenotype of acute pancreatitis disease from acute abdominal pain in the emergency setting (Chapter 3). A higher potential of NMR for non-target analysis in urine than for plasma samples in this particular application was shown. NMR was also able to measure the influence of drug treatments and therapeutic interventions on the pathological phenotype. Moreover, the metabolic differences in three clinical scenarios: *i*) diagnosis of the disease; *ii*) aetiology of acute pancreatitis (alcoholism and cholelithiasis); and *iii*) severity, were achieved after proper data treatment. Furthermore, separation inside the heterogeneous control group was proven between cholelithiasis and colon inflammation.
4. The integration combining profiles obtained by two complementary analytical techniques such as RMN and CE–UV (here with two separation modes profiles) was achieved (Chapter 4). The metabolic non–targeted analysis allowed the urinary characterization of a knockout mice model (slc26a6 deletion) due to the high capabilities of both techniques for this type of sample. Therefore, differences between the gender and the type were defined. Moreover, the identification of a metabolite in CE–UV through the correlation study with CE–NMR–SHY was obtained. In addition, metabolic relations of the two most significant metabolites (oxalate in CE–UV and m–HPPS on NMR) with other signals in the profile in the study were exposed.

In conjunction of this article (annex section), another one showed the CE–UV methodology as a tool for the urinary metabolic characterization of children with diabetes 1. Results from this study demonstrated the usefulness of this inexpensive and fast equipment for analysis in the evolution of the metabolic profile in diabetic children after 1 year with a nutraceutical diet enriched with antioxidants plus ω–3 PUFAs.

5. The major results obtained by LC–QTOF–MS demonstrated the high capabilities of this technique for finding distinct biochemical differences between patients who respond to treatment and those who do not (Chapter 5). This was the first study of the plasma metabolomic patterns in patients receiving ketamine for the treatment of bipolar disorder. Therefore, a metabolomics screening may be a guide for the prediction of the response. The appropriate data–treatment showed reliable results supported by reported literature.
6. The high selectivity and sensitivity properties of CE–LIF were employed for the determination of 14 amino acids in human plasma. The method was fully validated for 10 amino acids including D– and L– serine enantiomers and successfully accomplished all the validation parameters. Undiluted human plasma samples can be processed by ultrafiltration, derivatized, and analysed for amino acid profiles. The developed and optimized methodology used for the quantitation of amino acids in plasma samples from patients with bipolar disorder was in agreement with published literature (Chapter 6).
7. The discussion and confrontation of the analytical platforms applied in this thesis (GC–MS, NMR, CE–UV, LC–MS and CE–LIF) showed that they do not compete, but instead they form an integral group of tools capable of studying the metabolome in distinct biofluids and approaches of metabolomics. Continuous improvements in these techniques and in the data treatment obtained have succeeded in moving the field forward to face the current challenges associated with the high scope that metabolomics studies follow.

CONCLUSIONES

En resumen, este trabajo ha resaltado la importancia de las plataformas de análisis dentro del campo de la metabolómica. Aunque no existe actualmente una única técnica que cumpla con el objetivo de medir todos los metabolitos en un organismo, este trabajo demostró la forma en que cada técnica analítica contribuyó a resolver distintas cuestiones biológicas de acuerdo a sus capacidades. En este trabajo no sólo se realizó una descripción completa de las características (ventajas y limitaciones) de las técnicas analíticas más relevantes empleadas en la metabolómica, también se discutieron las diversas inquietudes y desafíos actuales relacionados con ellas. Más aún, para cada una de ellas, su empleo y el subsecuente tratamiento de los datos generado para llegar a la interpretación biológica, fue descrito.

La enorme importancia de estas técnicas analíticas; GC-MS, RMN, CE-UV, LC-MS y CE-LIF fue demostrada en las diferentes aplicaciones biológicas dentro de cada una de las estrategias de la metabolómica; el *desarrollo analítico de métodos y validación*, el *análisis metabolómico no-dirigido* y el *análisis dirigido*. Además, a través de estos capítulos de la evolución de las distintas técnicas analíticas utilizadas en la metabolómica puede ser seguida, desde equipos muy básicos (e.g., CE-UV), instrumentos de segunda generación (e.g., RMN y GC-MS) o de instrumentación más reciente (e.g., LC-MS). Más aún, la combinación o complementariedad entre ellas fue demostrada.

Las conclusiones particulares de este trabajo se enumeran por capítulos:

1. En un estudio a profundidad de la literatura acerca los métodos del análisis no-dirigidos en muestras biológicas por CE ya sea con detección UV o MS, permitió una evaluación crítica de las valiosas características de la CE en el análisis de los diferentes fluidos biológicos especialmente en muestras de orina debido a la abundancia de compuestos iónicos y polares. Entre sus puntos fuertes; la alta eficacia y resolución, el pequeño volumen de muestra y el mínimo tratamiento de muestra fueron los aspectos más significativos, mientras que su principal inconveniente fue su robustez (Capítulo 1).
2. Por primera vez, el metaboloma de la leche materna obtenido mediante un sistema de extracción en una única fase y caracterizado por dos técnicas instrumentales; la GC-Q-MS con alto potencial para el estudio de los metabolitos en muestras biológicas que pueden ser volátiles o térmicamente

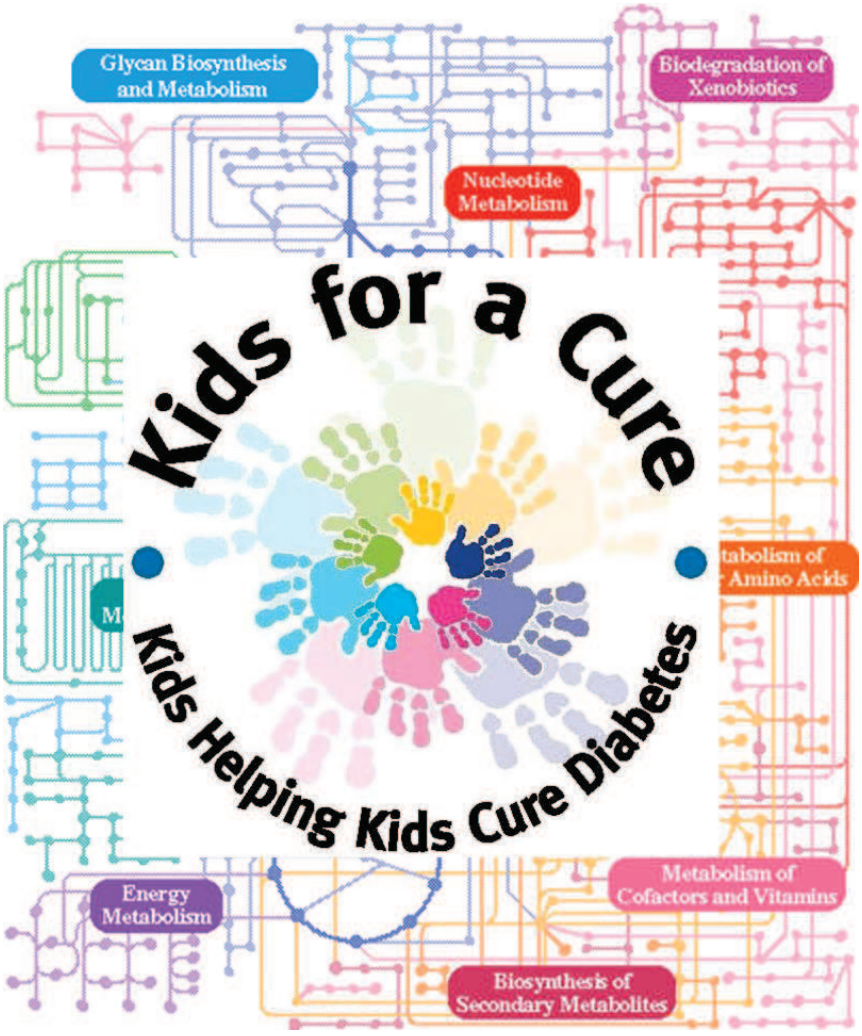
estables después de un proceso de derivatización y la LC-TOF-MS, donde diferentes y complementarias clases de compuestos bioquímicos fueron observados (Capítulo 2). Las metodologías en ambos instrumentos fueron completamente validadas presentando valores aceptables para estudios biológicos, y demostrando su valor para su aplicación en estudios posteriores. Por otra parte, la evolución de la leche materna en el tiempo fue establecida, confirmando los cambios previamente descritos. Aquí los ácidos grasos y el colesterol fueron significativos en ambos técnicas y siguieron el mismo patrón, los glicerolípidos y los fosfolípidos fueron propios de la LC-MS y, finalmente los hidratos de carbono de la GC-MS.

3. El uso del análisis no dirigido por RMN desplegó una estrategia de tratamiento de datos para determinar el fenotipo para el diagnóstico y pronóstico de la enfermedad de pancreatitis aguda del dolor abdominal agudo en el servicio de urgencias (Capítulo 3). En este estudio en particular se observó un mayor potencial de la RMN para el análisis no-dirigido en muestras de orina que en muestras de plasma. La medición por RMN permitió observar la influencia de los tratamientos farmacológicos y las intervenciones terapéuticas en el fenotipo patológico. Por otra parte, fueron establecidas las diferencias metabólicas en tres escenarios clínicos; *i*) en el diagnóstico de la enfermedad, *ii*) en la etiología de la pancreatitis aguda (alcoholismo y colelitiasis) y *iii*) en la gravedad. Además, se encontraron las diferencias dentro del grupo heterogéneo de los controles, distinguiendo entre pacientes con colelitiasis y con inflamación del colon.
4. La integración combinando los perfiles obtenidos mediante dos técnicas analíticas complementarias tales como RMN y CE-UV (esta utilizando dos modos de separación) fue conseguida (Capítulo 4). El análisis metabólico no-dirigido permitió la caracterización urinaria de un modelo de ratones knockout slc26a6, debido a las grandes capacidades de ambas técnicas para este tipo de muestra. Por lo tanto, las diferencias entre el sexo y el genotipo fueron definidas. Por otra parte, la identificación de un metabolito en CE-UV a través del estudio de correlaciones con RMN fue obtenida. Finalmente, las relaciones metabólicas de los dos metabolitos más significativos: del oxalato en CE-UV y el m-HPPS en RMN, con otras señales en el perfil global del estudio fueron expuestas.

   En conjunto con este artículo (en la sección del anexo), otro artículo de ejemplo mostró a la metodología en CE- UV como una herramienta valiosa para la caracterización metabólica de orina en niños con diabetes tipo 1. Los resultados de este estudio mostraron la utilidad de este

equipo de bajo costo y rápido para el análisis de la evolución del perfil metabólico en niños diabéticos después de estar 1 año con una dieta nutracéutica enriquecida en antioxidantes y PUFAs ω-3.

5. Los resultados más importantes obtenidos por LC-QTOF-MS, demostraron las altas capacidades de esta técnica para encontrar las diferencias bioquímicas entre los grupos de pacientes bipolares que respondieron al tratamiento con ketamina de los que no lo hicieron (Capítulo 5). Este fue el primer estudio de metabolómica en plasma de pacientes que recibieron ketamina para el tratamiento del trastorno bipolar. Por lo tanto, un cribado metabolómico puede ser una guía para la predicción de la respuesta. El tratamiento de datos apropiado mostró resultados fiables que fueron sustentados por la literatura reportada.
6. La alta selectividad y sensibilidad de CE-LIF, fueron empleadas para la determinación de 14 aminoácidos en el plasma humano. El método fue completamente validado para 10 aminoácidos incluyendo los enantiómeros D- y L-serina obteniendo con éxito todos los parámetros de validación. Las muestras sin diluir de plasma humano puede ser procesadas por ultrafiltración, derivatización y analizadas para obtener los perfiles de aminoácidos en CE-LIF. La metodología desarrollada y optimizada usada para la cuantificación de aminoácidos en muestras de plasma de pacientes con trastorno bipolar estuvo de acuerdo con los valores reportados en la literatura (Capítulo 6).
7. La discusión y confrontación de las plataformas analíticas aplicadas en esta tesis (GC-MS, RMN, CE-UV, LC-MS y CE-LIF) mostraron que estas no compiten, sino que por el contrario, forman un grupo integral de herramientas capaces de estudiar el metaboloma en los distintos fluidos biológicos y enfoques de la metabolómica. Las continuas mejoras en estas técnicas y en el tratamiento de los datos, han logrado avanzar para hacer frente a los retos actuales, asociados al alto alcance que los estudios de metabolómica persiguen.


Glycan Biosynthesis and Metabolism
Biodegradation of Xenobiotics
Nucleotide Metabolism
Kids for a Cure
Kids Helping Kids Cure Diabetes
tabolism of Amino Acids
Energy Metabolism
Metabolism of Cofactors and Vitamins
Biosynthesis of Secondary Metabolites

ANNEX

Journal of Pharmaceutical and Biomedical Analysis 53 (2010) 1298–1304

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Metabolomic approach to the nutraceutical effect of rosemary extract plus ω-3 PUFAs in diabetic children with capillary electrophoresis

Claudia Balderas[a,1], Alma Villaseñor[a,1], Antonia García[a], F. Javier Rupérez[a], Elena Ibañez[b], Javier Señorans[c], Julio Guerrero-Fernández[d], Isabel González-Casado[d], Ricardo Gracia-Bouthelier[d], Coral Barbas[a,*]

[a] CEMBIO (Center for Metabolomics and Bioanalysis) Facultad de Farmacia, Universidad San Pablo-CEU, Campus Montepríncipe, Boadilla del Monte, 28668 Madrid, Spain
[b] Instituto de Fermentaciones Industriales, CSIC, Madrid, Spain
[c] Sección Departamental de Ciencias de la Alimentación, Universidad Autónoma de Madrid, Madrid, Spain
[d] Servicio de Endocrinología Infantil, Hospital La Paz, Madrid, Spain

ARTICLE INFO

Article history:
Received 2 July 2010
Received in revised form 20 July 2010
Accepted 21 July 2010
Available online 30 July 2010

Keywords:
Metabolic fingerprint
Urine
Rosmarinus officinalis
Antioxidants
Diet

ABSTRACT

Type 1 diabetes mellitus is a major endocrine disorder, affecting approximately 5% of the world's population. It not only leads to hyperglycaemia but also causes many complications, and numerous studies have demonstrated that oxidative stress contributes to these complications. As a new strategy to improve the oxidative damage in diabetes, interest has grown in the usage of natural antioxidants, even more in the long term. Among them, *Rosmarinus officinalis* (rosemary) has been widely accepted as one of the species with the highest antioxidant activity. In addition, ω-3 polyunsaturated fatty acids were efficient in delaying and decreasing cardiovascular risk factors associated with diabetes.

Type 1 diabetic children and the corresponding controls were enrolled in the assay. The aim was evaluating the effect of a special additive containing rosemary extract, vitamin E and PUFAs added to their standard diet through the meat. In the analytical point of view, a metabolomic approach with CE-UV was used to detect possible differences in urine of diabetic children as compared to controls. After the application of the appropriate multivariate statistical tools, clear differences could be observed between treated and non-treated diabetic children and some of the metabolites associated could be identified. This was specially challenging as most of the clinical biochemical parameters measured by target analysis showed no differences between the groups.

1. Introduction

Serious long-term complications of diabetes include cardiovascular diseases, chronic renal failure, retinal damage [1,2] and most of them have been related to oxidative stress [3,4]. Moreover, they appear even under a good glycemic control [5].

In terms of diabetic's lipidic profile, it has been demonstrated that under insulin therapy it can became normal although some lipoproteic changes may remain thus perpetuating the cardiovascular risks [6].

Therefore, to be able to reduce the long-term effects of both, oxidative stress and lipoproteic change, an antioxidant therapy combined with a diet rich in ω-3 fatty acids has been suggested, that might act modulating different processes associated to diabetes disease [7].

As for the antioxidant therapy, numerous clinical trials have tried to demonstrate the benefits of a diet supplemented, for instance, with α-tocopherol, ascorbic acid or β-carotene in type 1 diabetes mellitus patients (reviewed in [8]). Nevertheless, no conclusive results were obtained in these studies in terms of the therapeutic effect of the diet in the treatment of long-term complications in type 1 diabetes.

In a previous paper we set up the scientific bases of new functional meat products with both a balanced ω-6/ω-3 ratio and a synergic combination of antioxidants such as supercritical rosemary extracts and vitamin E [9] and proved that they showed a lipidic profile closer to fresh salmon than other meat products and an antioxidant activity similar to fruit functional beverages. Thus, these meat products could be good candidates to test the efficacy of a combined diet containing antioxidants and ω-3 PUFAs.

What seems clear is that the metabolic control of glucose and lipids plays a significant role in the prevention of the long-term complications of type 1 diabetes [10] but more information is required to decipher the real effects of an antioxidant therapy as

* Corresponding author. Tel.: +34 91 3724711; fax: +34 91 3724712.
E-mail address: cbarbas@ceu.es (C. Barbas).
[1] Equal work in the manuscript.

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.07.034

a way to prevent or retard the appearance of symptoms related to oxidative stress in diabetic patients.

In order to accurately follow the 'therapy' for a chronic disease such as type 1 diabetes, normalization of a targeted aspect of the metabolism must occur (without disruption of other metabolic pathway regulation). Furthermore, it is increasingly recognized that assessment of limited biomarker compounds to monitor therapy efficacy is fundamentally flawed and that more comprehensive snapshots of multiple metabolites must be taken. Such an approach is commonly employed in metabolomics investigations. Metabolic fingerprinting is a complex matrix profiling strategy widely adopted by many researchers and which can be applied to a variety of sample matrices [11,12]. Metabol(n)omics has gained great prominence in diabetes research within the last few years and has already been applied to understand the metabolism in a range of animal models and, more recently, attempts have been done to process complex metabolic data sets from clinical studies. A recent review summarizes the technologies currently being used in metabol(n)omics, as well as the studies reported related to diabetes [13].

The latter member of the metabolomics analytical toolbox—CE is proving a truly useful addition, especially given its high-throughput potential and low sample volume requirement. Another significant benefit of CE for metabolomics studies includes the ability to measure all type of analytes in a sample (which is only limited by the detection system) *via* employment of multiple separation mechanism modes on a single sample [14–19].

The aim of this study was checking the capabilities of CE-UV to detect differences in urine of diabetic children as compared to controls and, in such case, studying the possible effect of a change in the diet of diabetic children, including designed meat products with an ω-6/ω-3 ratio lower than 4 and a combination of natural antioxidants (supercritical rosemary extracts and vitamin E) at controlled dosage for 1 year.

2. Materials and methods

2.1. Chemicals

Sodium tetraborate decahydrate (STD), β-cyclodextrin sulphated (analytical grade, Sβ-CD) and methanol were purchased from Sigma–Aldrich (Steinheim, Germany), sodium dodecyl sulphate (SDS) and sodium hydroxide from Panreac Química S.A.U. (Barcelona, Spain), and hydrochloric acid from Fluka (Buchs, Switzerland). Standards used for peak identification were obtained from Sigma, except 2-OH-butyric acid, 3-OH-butyric acid, glutaric acid, guanine, L-pyroglutamic acid, DL-serine, 2,3-dihydroxybenzoic acid from Fluka, L-tyrosine from Merck (Darmstadt, Germany), sodium oxalate from Panreac. Reverse osmosed deionised water (Milli-Q Synthesis from Millipore, Bedford, MA, USA) was used for standard solution and electrolyte preparations.

2.2. Study design

33 type 1 diabetic children and 16 control ranging from 6 to 11 years old for boys and girls and including 27 of boys and 22 girls were enrolled in the assay. All parents signed informed-consent forms approved by the ethics committee of La Paz Hospital. This was an exploratory double blind randomized study.

The aim was evaluating the effect of a special extract containing 0.02% rosemary extract, 0.001% vitamin E and 0.3% PUFAs on T1DM young children when it was added to meat products (100 g per day, 3 days a week) as part of their regular diet. The functional meat product was supplied by Grupo Frial S.A. (Tres Cantos, Spain).

Children were divided into three groups: diabetic children with and without the extract and control children with the extract. All of them received the same meat products prepared weekly by a local industry (Grupo Frial S.A.). The diet was administered during 12 months and in each visit the adhesion was assessed. Mean adhesion to diet was around 80% and nobody was under 60%.

2.3. Urine samples

Punctual urine samples were collected at time zero, when the study started and after 12 months. Samples were kept at −80 °C until the analysis. The day of analysis the urine was defrosted and aliquoted for each CE method, vortexed and for CZE method 8 μl of urine were diluted with 32 μl of deionised water, for CD-MEKC method 10 μl urine were mixed with 10 μl BGE1 and 80 μl deionised water, and injected directly into the CE apparatus. All samples were initially run with negative polarity and then with positive polarity, using a separated aliquot for each method.

2.4. Instrumentation

CE experiments were carried out on a P/ACE MDQ system (Beckman Coulter, Fullerton, CA, USA) equipped with diode array UV-absorbance detection (190–600 nm), a temperature-controlled (liquid cooled) capillary compartment and an autosampler. Electrophoretic data were acquired and analysed with 32 Karats software (P/ACE MDQ instrument). Separations achieved with normal polarity were performed in a fused silica capillary (75 μm internal diameter), 60 cm total, 50 cm effective length (Beckman Coulter). New capillaries were conditioned for 5 min at 25 °C with water, followed by 0.1 M NaOH for 15 min, 8 min with deionised water and finally BGE(1) for 10 min. Before each analysis, the capillaries were washed with 0.1 M HCl and deionised water for 5 min, and then 5 min with the run buffer. The running buffer (BGE1) comprised 25 mM STD, 75 mM SDS and 1.43% (p/v) β-cyclodextrin sulphated. The pH was adjusted to pH 9.50 with 2 M NaOH. The capillaries were maintained at 20 °C, with 20 kV applied voltage and 10 s hydrodynamic injection (0.5 psi).

Separations achieved with reverse polarity were performed using a capillary (Beckman Coulter) coated with polyacrylamide 60 cm of total length, 50 cm effective length and 50 μm internal diameter. On first use, the capillaries were conditioned by a pressure flush of 0.1 M HCl (2 min) and BGE (10 min). Between runs, the capillaries were flushed under pressure with deionised water (2 min) and BGE (5 min). All experiments were performed at 25 °C using a separation potential of −20 kV. Samples were injected at the cathode, with 0.5 psi of pressure applied for 25 s. Resolved sample components were detected at the anode. Electrophoretic buffer (pH 6.10) was prepared with 0.2 M ortho-phosphoric acid, adjusted to pH with NaOH saturated, and 10% (v/v) methanol HPLC grade was added. (The current observed under these conditions was −86 to −90 μA.) Data were collected at a frequency of 4 Hz. Buffer solutions were filtered through a 0.2 μm nylon filter before use.

2.5. Data alignment and treatment

Baseline correction and multialignments of the electropherograms were performed with an in-house program developed in Matlab® 7.0. For baseline correction the profiles were divided in p zones, in each zone the minimum value was accounted and all the minima were linearly interpolated. Multialignment was performed using the Correlation Optimized Warping (COW) method previously described [20]. For normalization, two approaches were tested: (1) the division of all signals by the creatinine peak (creatinine is often used to correct urine dilutions effects due to its

Table 1
Clinical parameters measured in children by La Paz Hospital at time zero.

Parameter		Glu** mg/dL	HbA1c** %	Total-Chol mg/dL	Chol-HDL mg/dL	TAG mg/dL	U. A.* mg/dL	TP* mg/dL	Creat mg/dL	mAlb mg/dL	Vit-E mg/dL
Patient											
Healthy ($n=16$)	Mean	81.1	5.2	179.4	64.4	56.3	3.7	7.2	0.7	4.3	1035.8
	STD	6.2	0.3	34.4	13.3	19.1	0.7	0.4	0.1	3.0	220.8
	RSD (%)	7.7	4.8	19.2	20.6	33.9	18.0	5.6	10.8	69.5	21.3
Diabetic ($n=34$)	Mean	172.9	7.6	172.0	63.1	50.0	3.1	6.9	0.7	4.0	1029.7
	STD	80.1	0.9	35.0	13.8	11.0	0.6	0.4	0.1	3.0	221.2
	RSD (%)	46.3	12.0	20.4	21.9	21.9	19.7	5.4	10.1	75.4	21.5

Glu, glucose; HbA1c, glycosylated haemoglobin; Total-Chol, total cholesterol; Chol-HDL, cholesterol HDL; TAG, triglycerides; U.A., uric acid; TP, total proteins; Creat, creatinine; mAlb, microalbumin; Vit. E, vitamin E.
* P value < 0.05.
** P value < 0.01.

excretion is almost constant during the day [21]). (2) Probabilistic quotient normalization. This method is based on the calculation of a most probable dilution factor by looking at the distribution of the quotients of the amplitudes of a test spectrum by those of a reference spectrum. It is demonstrated to be more robust and more accurate than other methods commonly employed [22]. Subsequently data were analysed with SIMCA P+ 12.0.1 (Umetrics, Sweden).

3. Results and discussion

Capillary electrophoresis provides a comprehensive snapshot of multiple metabolites in biological samples especially in urine because all analytes are already dissolved and most of them are easily separated due to its charge; despite its detection system, CE gives a general metabolic response. With different CE modes we can obtain a complete profile of a wide set of compounds charged and neutral using both polarities producing an extended representation [14,17–19,21,23,24]. The two methods operated with opposite polarities and orthogonal capillary electrophoretic separation modes. Method 1: normal polarity cyclodextrin modified micellar electrokinetic chromatography (CD-MEKC), mostly for cations and neutral compounds, and method 2: reversed polarity capillary zone electrophoresis (CZE), mainly for anions. The methodology already proved to be successful with urine from control and diabetic rats [14], but in humans with different diet, physiological, environment and genetic variation are still a challenge.

Several clinical and biochemical parameters (glucose, glycosylated haemoglobin (HbA1c), total cholesterol, HDL-cholesterol, total proteins, triglycerides, creatinine, uric acid, vitamin E and microalbumin) were measured in these children at the beginning of the study and no differences were found except for glucose and HbA1, which were higher in diabetics while uric acid and total proteins were significantly lower (Table 1). Therefore the first challenge was to test the capability of our methodology to discriminate between the two groups, considering that none of these parameters were detected.

Fingerprints were obtained with a relatively inexpensive tool such as CE-UV, without any previous treatment other than filtration. After that, data pre-treatment of electrophoretic profiles (baseline correction, alignment, scaling and normalization) was applied. That step proved to be critical to obtain satisfactory results as CE profiles usually are affected by some type of shift, and different natural degrees of dilution in urine can alter the comparison. The fingerprints obtained with two polarities after alignment are shown in Fig. 1.

Normalization is a pre-processing method which accounts for different dilutions of samples by scaling the spectra to the same virtual overall concentration. Best results, considering sample clustering and quality of the models, were obtained with the probabilistic quotient normalization and that was the procedure used throughout the work.

Finally chemometric tools such as PCA (principal components analysis), PLS-DA (partial least squares discriminant analysis) and OPLS-DA (orthogonal partial least squares discriminant analysis) were applied. Regarding multivariate data analysis, a satisfactory result for pattern recognition was obtained with a supervised analysis such as OPLS-DA using Pareto scaling (most commonly used for metabolomic research), as can be observed in Fig. 2, after deleting from the group several strong outliers, according to the Hotellings T^2 range plot (SIMCA Umetrics).

Fingerprinting ignores the assignment problem presented by the multitude of signals and, instead, uses multivariate analysis to compare sets of profiles and hence the samples from which the profiles were derived. Nevertheless, once differences between patterns have been established, the obvious interest is the identification of as many peaks in the profile as possible in order to obtain

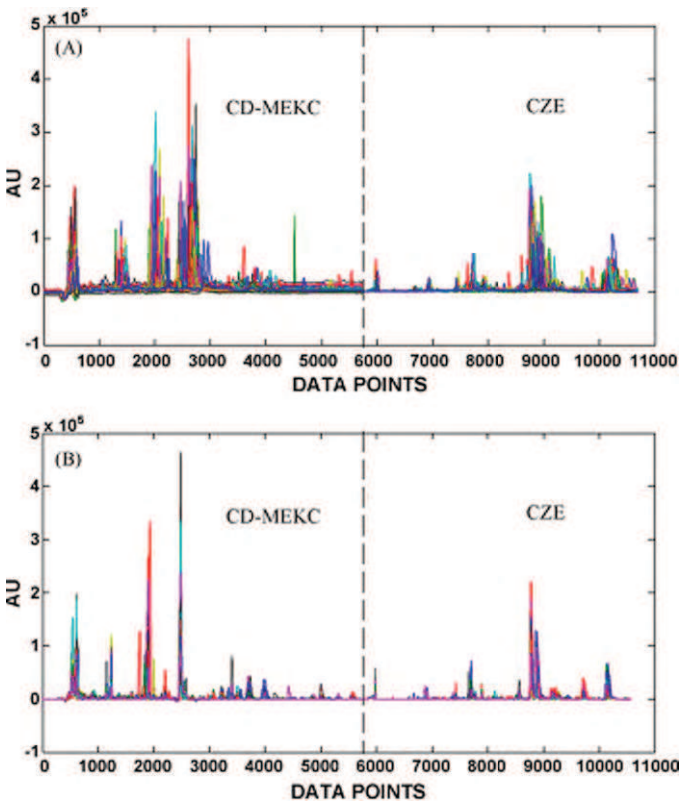


Fig. 1. (A) Raw data with the results coming from both polarities added up (first half normal polarity, second half second polarity) and (B) 98 profiles (CE-UV) after alignment and scaling with COW algorithm after optimizing the slack.

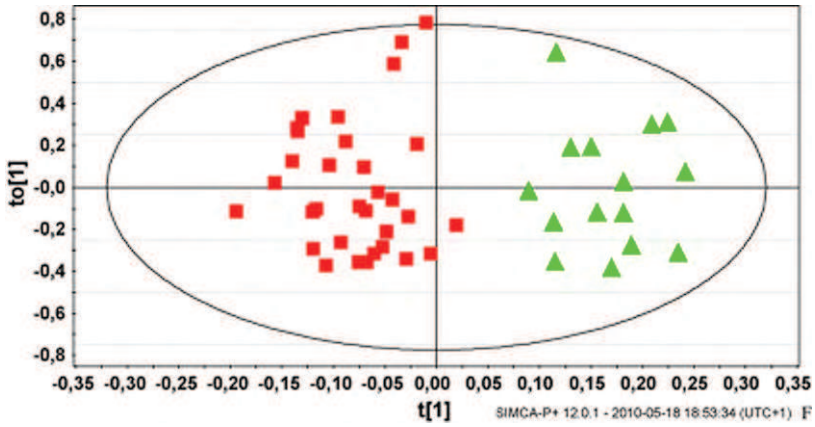


Fig. 2. OPLS-DA data derived using the total profile added of CD-MEKC (normal polarity) and CZE (reverse polarity) methods at zero months. Identification: ▲, healthy; ■, diabetic.

biochemical insight into the pathology condition. This is possibly the weakest point in CE with UV detection, because identification can only be made *via* trial and error, using pure standards of compounds expected in the profile. The assignment is performed by comparison of migration times and spectra and by spiking the sample with the pure standards. Fig. 3 displays 26 metabolites using CD-MEKC method (normal polarity) from a total of 71 compounds tested; some of the identified standards gave analytical signal in the same migration time and increased the same peak into the profile. Fig. 4 shows an example of an electropherogram of a urine sample in which 30 compounds were identified by CZE method (reverse polarity) out of 76 standards tested.

After that, the corresponding loading plots were employed to identify those compounds that contributed most to the sample grouping. Among the compounds identified so far, we noticed increased nitrites, citrate, phenyllactate, glutamate, creatinine and urea in diabetic children and lower quantities of glutarate, guanidine, phospho-L-serine, benzoate, urate, and glycerate (Table 2). In addition, there were 14 significant metabolites still unknown and for that reason they have not been included in the table.

In a previous study [25] with Proton Nuclear Magnetic Resonance (^{1}H NMR), profiles of urine from children and adolescents with type 1 diabetes showed highly significant differences for citrate, alanine, lactate, and hippurate, which were all higher in diabetic children than in the unaffected individuals. According to these authors, the consistently higher citrate, alanine, and hippurate might reflect the increased glomerular filtration rate characteristic of type 1 diabetes [26] and/or a modification of the transport mechanisms at the tubular level. The latter may be related either to altered cellular function or to the presence of high glucose concentrations in the tubular lumen. It should be also taken into account

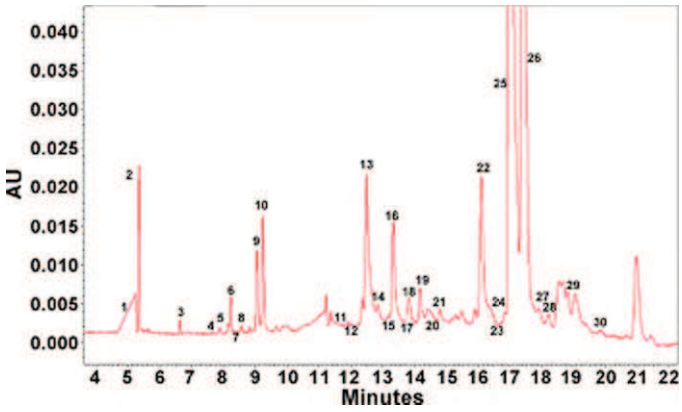


Fig. 4. Electropherogram profile for urinary sample obtained using the reverse polarity CZE method (for conditions see text). Peak identification: 1, chloride; 2, nitrate; 3, oxalate; 4, fumarate; 5, trans-aconitate; 6, 2-ketoglutarate; 7, malate/succinate; 8, isocitrate; 9, glutarate; 10, citrate; 11, acetoacetate; 12, lactate; 13, O-phospho-L-serine; 14, glycerate; 15, 2-hydroxybutyrate; 16, benzoate; 17, 3-hydroxybutyrate; 18, pyroglutamate; 19, 2,5-dihydroxybenzoate/glutamine; 20, 2,3-dihydroxybenzoate; 21, aspartate; 22, β-phenyllactate; 23, quinoline-2-carboxylate; 24, glutamate; 25, hippurate; 26, urate; 27, p-hydroxyphenyllactate; 28, homovanillate; 29, vanillylmandelate/2-aminoadipate; 30, glucuronate.

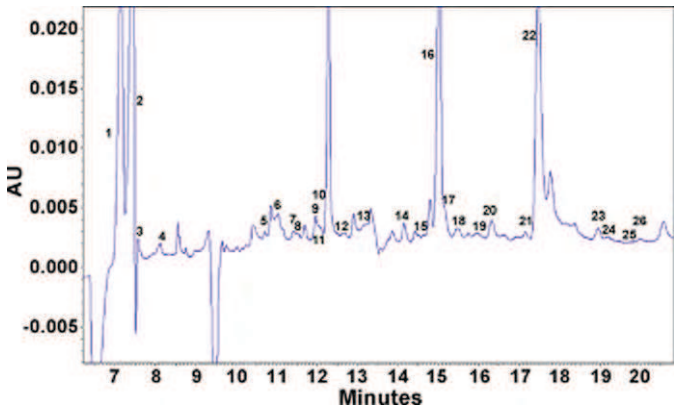


Fig. 3. Electropherogram profile for urinary sample obtained using the normal polarity CD-MEKC method (for conditions see text). Peak identification: 1, urea; 2, creatinine; 3, guanidinoacetate acid; 4, ornithine; 5, histidine; 6, tyrosine, 7, metionine; 8, serine; 9, phenylalanine; 10, 5-methyluridine; 11, 2,5-dihydroxybenzoate/glutamine; 12, tryptophan; 13, asparagine; 14, phenaceturate; 15, uridine; 16, hippurate; 17, β-phenyllactate; 18, guanosine; 19, p-hydroxyphenyllactate; 20, quinaldate; 21, inosine; 22, urate; 23, pyroglutamate; 24, β-3,4-dihydroxyphenylalanine; 25, isovanillate; 26, 3-nitro-L-tyrosine.

Table 2
Discriminant metabolites found using OPLS-DA model for type 1 diabetic disease in children before the treatment as compared to controls.

Metabolites	Diabetics
Nitrite	↑
Citrate	↑
Glutarate	↓
Phenyllactate	↑
Glutamate	↑
Creatinine	↑
Guanidinoacetate	↓
Phospho-L-serine	↓
Benzoate	↓
Urea	↑
Urate	↓
Glycerate	↓

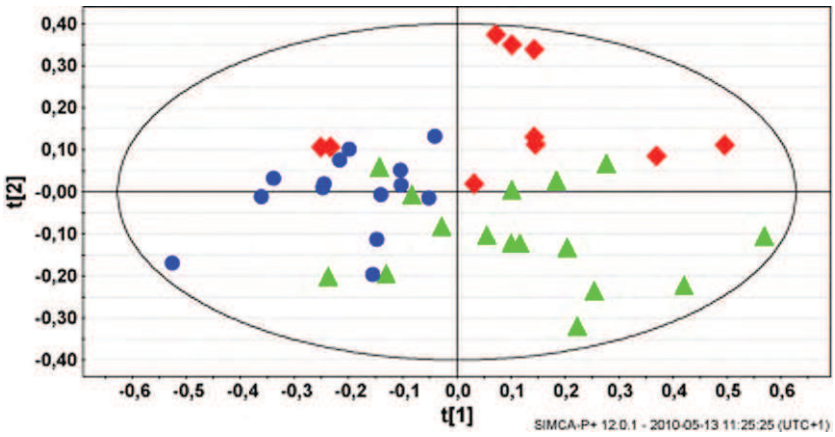


Fig. 5. PLS-DA data derived using the total profile added of CD-MEKC (normal polarity) and CZE (reverse polarity) methods at 12 months. Identification: ▲, healthy children with extract (H); ♦, diabetic children treated with extract (TD) and ●, diabetic children non-treated with the extract (NTD).

that activity of ATP-citrate lyase, a cytosolic enzyme, is enhanced by insulin; therefore accumulation of citrate may occur in absence of insulin and/or insulin resistance.

A different study with Doppler ultrasonography [27] found urinary nitrite and nitrate concentrations significantly increased in children with diabetes compared with control subjects. In particular, prepubertal children with diabetes had significantly higher $NO_2^- + NO_3^-$ serum levels than prepubertal healthy children. The authors concluded that this study demonstrates that in children with diabetes, chronic hyperglycaemia may act through a mechanism that involves increased NO production and/or action and contributes to generating intrarenal hemodynamic abnormalities [28].

Urines obtained after 12-month treatment were submitted to the same procedure and PLS-DA results are shown in Fig. 5. Quality parameters of the model showed a good result for the explained variance ($R^2 = 0.64$), but a poor one for the predicted variance ($Q^2 = -0.16$).

Comparing the three groups, results showed that there is a significant change in metabolism due to the diet. In Fig. 5 it can be appreciated that latent variable 1 seems to be related to the diet while variable 2 is more related to the pathology. For analysing metabolites playing a more significant role, we compared in OPLS-DA each two groups making three different comparisons that are showed in Fig. 6 for the scores plot. In addition we obtained the loadings S-plot where each variable is represented by its contribution in the group (covariance) and its confidence (correlation). The interval of covariance selected was ±0.15 and those variables which resulted significant are summarized in Table 3. In addition, Jack-knife intervals were obtained to check homogeneities, also included in the table.

Guanidinoacetate and creatinine are metabolites related to catabolism of aminoacids (glycine) that is known to be increased in diabetics due to impairment of insulin. Although treatment has not decreased this catabolism, the accumulation of guanidinoac-

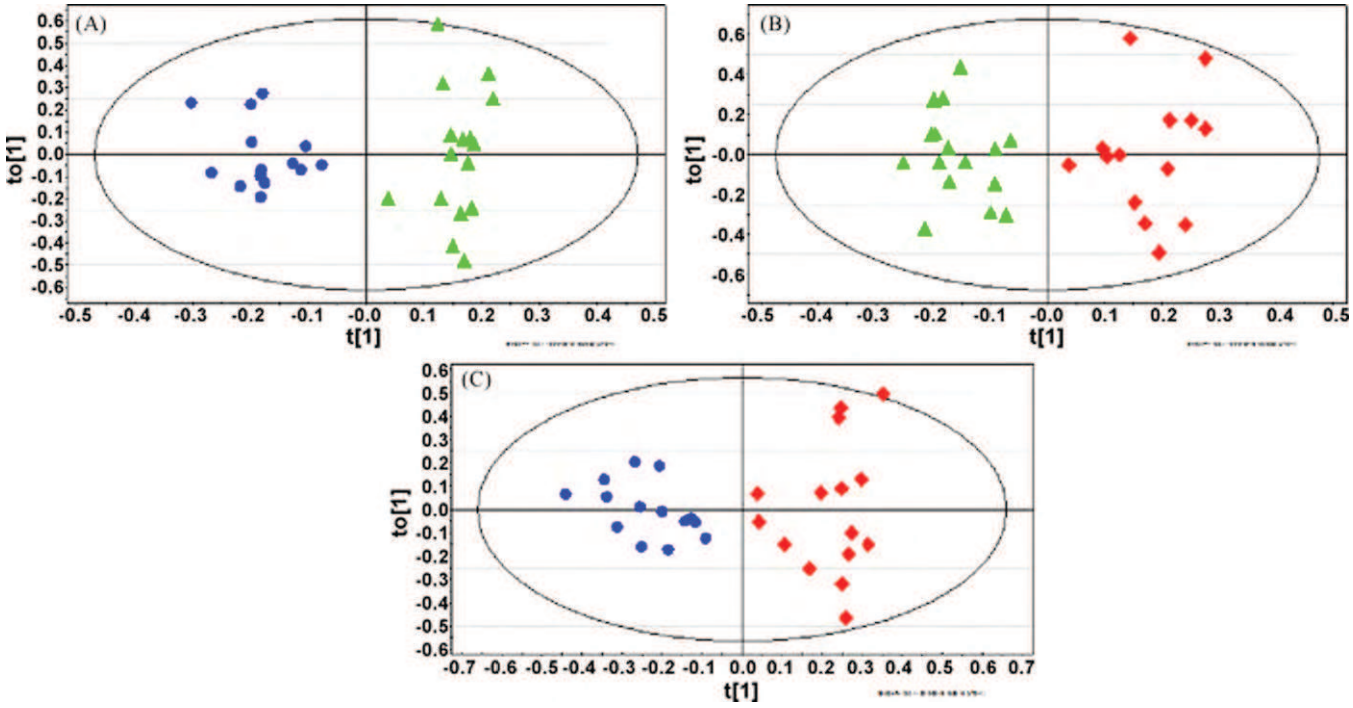


Fig. 6. Different comparisons of OPLS-DA data derived using the total profile obtained by adding CD-MEKC (normal polarity) and CZE (reverse polarity) methods at 12 months. (A) Healthy with extract vs. diabetic children without extract; (B) healthy vs. diabetic children both with extract; (C) diabetic children with and without extract. Identification: ▲, healthy children with extract; ♦, diabetic children with extract and ●, diabetic children without extract.

Table 3
Discriminant metabolites found in OPLS-DA model for type 1 diabetic disease in children after 12 months of receiving the extract. (i) Treated diabetic (TD) vs. healthy children (H); (ii) non-treated diabetic (NTD) vs. healthy children (H) and (iii) non-treated diabetic (NTD) vs. treated diabetic (TD). Homogeneity was expressed with letters: common letter(s) mean no evidence enough for significant differences, being *a* the letter corresponding to the lowest value.

Metabolites	Comparisons			Homogeneity		
	TD/H	NTD/H	NTD/TD	H	NTD	TD
Nitrite	↑	↓	↓	b	a	c
Citrate	↑	↑	↓	a	b	c
Ketoglutarate	↑	↑		a	a	b
Aminoadipate	↑	↓	↓	b	a	c
Phenyllactate	↑			a	ab	b
Glutamate	↑	↑		a	a	b
Creatinine	↑		↓	a	a	b
Guanidinocetate		↑	↑	a	b	a
Phospho-L-serine	↑		↓	a	a	b
Pyroglutamate	↑			a	ab	b
Benzoate		↓	↓	b	a	ab
Urea	↓	↑	↑	b	c	a
Urate		↓	↓	b	a	b
3-Hydroxybutyrate		↑		a	b	ab
p-Hydroxyphenyllactate	↓	↓	↓	c	a	b
Phenaceturate			↓	ab	a	b

11 significant metabolites were unknown and are not assigned in the table.

etate seen in diabetic children without diet was overcome and catabolism could proceed to the last step when they received the supplemented diet. Urea changes must be investigated more in detail: although the reduction in the increased urea excretion characteristic of diabetics should be considered positive, the level of controls was found to be higher than in treated diabetics, and maybe not all the ammonium was excreted as urea. Increase in aminoadipate should be treated with caution as well, because the level of this metabolite of tryptophan, lysine and OH-lysine catabolism when diabetes and diet were both present was higher than in controls. Urate has long been considered the main antioxidant of non-lipidic media, and reductions of it have been attributed to oxidative stress increase [29]. When children took the diet with adequate PUFAs and antioxidants they presented urate levels not different from healthy controls. In addition, 3-OH butyrate in treated diabetic was not different from controls, pointing at a better metabolic control (less ketogenesis). Insulin deficiency with glucagon excess leads to the release of ketone bodies (acetoacetate, 3-OH-butyrate) by the liver and excretion in the urine [30].

It is also noteworthy that some metabolites with aromatic ring, such as phenyllactate and p-OH-phenyllactate, phenaceturate and benzoate have been found to be significantly changed due to diabetes and/or treatment. Their detection is related to the analytical tools used (UV) but they have not been reported previously as altered in this condition, and probably deserve further studies.

Results presented demonstrated the usefulness of CE with UV detection for metabolomics. The use of this inexpensive and fast tool allowed following the evolution of the metabolic profile of diabetic children with a diet modified by the consumption of meat products enriched with antioxidants (supercritical rosemary extract plus vitamin E) plus ω-3 PUFAs compared to diabetic children with no enriched meat products and with healthy children. After the application of the appropriate statistical tools, clear differences could be observed between treated and non-treated diabetic children and some of the metabolites associated could be identified. More in depth analytical and biochemical interpretation of the results is needed to understand the real effect of the diet in the antioxidant status of the diabetics. Nevertheless, this is an encouraging starting point to apply other technologies with higher elucidation capabilities.

Acknowledgements

Alma Villaseñor and Claudia Balderas acknowledge EADS-CASA for their fellowships. The authors gratefully acknowledge the financial support from Grupo Frial S.A. (Project CENIT HIGEA CEN-20072003) and Ministry of Science and Technology (MCIT) CTQ2008-03779.

References

[1] H. Bangstad, I. Seljeflot, T. Berg, K. Hanssen, Renal tubulointerstitial expansion is associated with endothelial dysfunction and inflammation in type 1 diabetes, Scand. J. Clin. Lab. Invest. 69 (2009) 138–144.
[2] C. Georgakopoulos, M. Eliopoulou, A. Exarchou, V. Tzimis, N. Pharmakakis, B. Spiliotis, Decreased contrast sensitivity in children and adolescents with type 1 diabetes mellitus, J. Pediatr. Ophthalmol. Strabismus (2010) 1–6.
[3] P. Pacher, I. Obrosova, J. Mabley, C. Szabó, Role of nitrosative stress and peroxynitrite in the pathogenesis of diabetic complications. Emerging new therapeutical strategies, Curr. Med. Chem. 12 (2005) 267–275.
[4] M. Goodarzi, A. Navidi, M. Rezaei, H. Babahmadi-Rezaei, Oxidative damage to DNA and lipids: correlation with protein glycation in patients with type 1 diabetes, J. Clin. Lab. Anal. 24 (2010) 72–76.
[5] The Diabetes Control and Complications Trial Research Group, The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, N. Engl. J. Med. 329 (1993) 977–986.
[6] B. Vergès, Dyslipidaemia in diabetes mellitus. Review of the main lipoprotein abnormalities and their consequences on the development of atherogenesis, Diabetes Metab. 25 (1999) 32–40.
[7] C. Cárdenas, E. Bordiu, J. Bagazgoitia, A. Calle-Pascual, S.D.A. Diabetes and Nutrition Study Group, Polyunsaturated fatty acid consumption may play a role in the onset and regression of microalbuminuria in well-controlled type 1 and type 2 diabetic people: a 7-year, prospective, population-based, observational multicenter study, Diabetes Care 27 (2004) 1454–1457.
[8] G. Caimi, C. Carollo, R. Lo Presti, Diabetes mellitus: oxidative stress and wine, Curr. Med. Res. Opin. 19 (2003) 581–586.
[9] G. Reglero, P. Frial, A. Cifuentes, M. García-Risco, L. Jaime, F. Marin, V. Palanca, A. Ruiz-Rodríguez, S. Santoyo, F. Señoráns, C. Soler-Rivas, C. Torres, E. Ibañez, Meat-based functional foods for dietary equilibrium omega-6/omega-3, Mol. Nutr. Food Res. 52 (2008) 1153–1161.
[10] WHO, in: WHO (Ed.), Diet, Nutrition and the Prevention of Chronic Diseases, WHO, Geneva, 2003, pp. 1–149.
[11] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics—a review in human disease diagnosis, Anal. Chim. Acta 659 (2010) 23–33.
[12] J. Sébédio, E. Pujos-Guillot, M. Ferrara, Metabolomics in evaluation of glucose disorders, Curr. Opin. Clin. Nutr. Metab. Care 12 (2009) 412–418.
[13] J. Faber, D. Malmodin, H. Toft, A. Maher, D. Crockford, E. Holmes, J. Nicholson, M. Dumas, D. Baunsgaard, Metabonomics in diabetes research, J. Diabetes Sci. Technol. 1 (2007) 549–557.
[14] M. Vallejo, S. Angulo, D. García-Martínez, A. García, C. Barbas, New perspective of diabetes response to an antioxidant treatment through metabolic fingerprinting of urine by capillary electrophoresis, J. Chromatogr. A 1187 (2008) 267–274.
[15] T. Soga, R. Baran, M. Suematsu, Y. Ueno, S. Ikeda, T. Sakurakawa, Y. Kakazu, T. Ishikawa, M. Robert, T. Nishioka, M. Tomita, Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption, J. Biol. Chem. 281 (2006) 16768–16776.
[16] R. Ramautar, O. Mayboroda, A. Deelder, G. Somsen, G. de Jong, Metabolic analysis of body fluids by capillary electrophoresis using noncovalently coated capillaries, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 370–374.
[17] C. Barbas, M. Vallejo, A. García, D. Barlow, M. Hanna-Brown, Capillary electrophoresis as a metabolomic tool in antioxidant therapy studies, J. Pharm. Biomed. Anal. 47 (2008) 388–398.
[18] I. Garcia-Perez, A. Couto Alves, S. Angulo, J. Li, J. Utzinger, T. Ebbels, C. Legido-Quigley, J. Nicholson, E. Holmes, C. Barbas, Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating Schistosoma mansoni infection in a mouse model, Anal. Chem. 82 (2010) 203–210.
[19] M. Vallejo, A. García, J. Tuñón, D. García-Martínez, S. Angulo, J. Martin-Ventura, L. Blanco-Colio, P. Almeida, J. Egido, C. Barbas, Plasma fingerprinting with GC–MS in acute coronary syndrome, Anal. Bioanal. Chem. 394 (2009) 1517–1524.
[20] G. Tomasi, F. van den Bergand, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom. 18 (2004) 231–241.
[21] S. Angulo, I. García-Pérez, C. Legido-Quigley, C. Barbas, The autocorrelation matrix probing biochemical relationships after metabolic fingerprinting with CE, Electrophoresis 30 (2009) 1221–1227.
[22] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics, Anal. Chem. 78 (2006) 4281–4290.

[23] I. García-Pérez, M. Vallejo, A. García, C. Legido-Quigley, C. Barbas, Metabolic fingerprinting with capillary electrophoresis, J. Chromatogr. A 1204 (2008) 130–139.
[24] F. Ruperez, D. Garcia-Martinez, B. Baena, N. Maeso, M. Vallejo, S. Angulo, A. Garcia, E. Ibañez, F. Señorans, A. Cifuentes, C. Barbas, Dunaliella salina extract effect on diabetic rats: metabolic fingerprinting and target metabolite analysis, J. Pharm. Biomed. Anal. 49 (2009) 786–792.
[25] C. Zuppi, I. Messana, P. Tapanainen, M. Knip, F. Vincenzoni, B. Giardina, M. Nuutinen, Proton nuclear magnetic resonance spectral profiles of urine from children and adolescents with type 1 diabetes, Clin. Chem. 48 (2002) 660–662.
[26] T. Costacou, R. Ferrell, D. Ellis, T. Orchard, Haptoglobin genotype and renal function decline in type 1 diabetes, Diabetes 58 (2009) 2904–2909.
[27] A. Savino, P. Pelliccia, C. Schiavone, A. Primavera, S. Tumini, A. Mohn, F. Chiarelli, Serum and urinary nitrites and nitrates and Doppler sonography in children with diabetes, Diabetes Care 29 (2006) 2676–2681.
[28] P. Cortes, F. Dumler, K. Venkatachalam, J. Goldman, K. Sastry, H. Venkatachalam, J. Bernstein, N. Levin, Alterations in glomerular RNA in diabetic rats: roles of glucagon and insulin, Kidney Int. 20 (1981) 491–499.
[29] G. Marra, P. Cotroneo, D. Pitocco, A. Manto, M. Di Leo, V. Ruotolo, S. Caputo, B. Giardina, G. Ghirlanda, S. Santini, Early increase of oxidative stress and reduced antioxidant defenses in patients with uncomplicated type 1 diabetes: a case for gender difference, Diabetes Care 25 (2002) 370–375.
[30] F. Prisco, A. Picardi, D. Iafusco, R. Lorini, L. Minicucci, M. Martinucci, S. Toni, F. Cerutti, I. Rabbone, R. Buzzetti, A. Crino, P. Pozzilli, Blood ketone bodies in patients with recent-onset type 1 diabetes (a multicenter study), Pediatr. Diabetes 7 (2006) 223–228.

GLOSSARY

Metabolomics related definitions.

BIOMARKER. This is a substance found in biological samples that is detected in higher (or lower) concentration than normal in patients with a certain disease[1].

COMBINATORIAL METABOLISM. The highly complex process whereby multiple independent metabolic systems (for example, in both host and microorganism) sequentially metabolize a compound, which results in a large number of possible metabolic products with diverse structures[2].

CO–METABOLITE. A metabolite that can only be formed by the integrated biochemical actions of more than one genome, e.g. the gut microbial metabolism of a mammalian metabolite or vice versa[2–3].

CO–METABOLOME. Set of compounds derived from interactions of more than one genome in symbiotic systems[4].

CONDITIONAL METABOLIC PHENOTYPE. A characteristic metabolite profile reflecting the host genome and its interaction with environmental factors, diet, and the gut microbiome[3–4].

ENDOGENOUS METABOLISM. Metabolic conversions under direct host cell genome/proteome control or under mitochondrial control, for example, all major energy–generating pathways and biosynthetic routes[5].

LIPIDOMICS. Chemical characterization of distinct lipid species in cells and the molecular mechanisms through which they facilitate cellular function[6].

METABOLIC NON–TARGET ANALYSIS. Unbiased, global screening approach to classify samples based on metabolite patterns or profiles that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites[7].

METABOLIC FOOTPRINTING. Non–target analysis of extra–cellular metabolites in cell culture medium as a reflection of metabolite excretion or uptake by cells[8].

METABOLIC PROFILING. Quantitative analysis of set of metabolites in a selected biochemical pathway or a specific class of compounds. This includes target analysis, the analysis of a very limited number of metabolites, e.g. single analytes as precursors or products of biochemical reactions[7].

METABOLIC TRAJECTORY. The visualization of the dynamic change in the multi– variate metabolite profile of a biofluid or tissue with time, as projected using chemometric methods[2].

METABOLITE. Small molecules that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell[8].

METABOLOME. The quantitative description of all endogenous low molecular weight molecules (<1 kDa) present in a biological sample[2-5, 8].

METABOLOMICS. The comprehensive quantitative analysis of all the metabolites of an organism or specified biological sample[3, 5, 7-9].

METABONOME. Theoretical combinations, sums, and products of the interactions of multiple metabolomes (genome, symbiotic, parasitic, environmental, and co–metabolic) in a complex system[3-5].

METABONOMICS. The quantitative measurement of the multivariate metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification[9, 10]. It is an approach to understanding the global metabolic regulation of an organism and its commensal and symbiotic partners[11]. The word origin is from the Greek (*meta* meaning change and *nomos* meaning a rule set or set of laws)[3–5].

METABOTYPE. A metabolic profile that defines a phenotype which relates to genetic variation of the mammal[2-3, 12].

MICROBIOME. The consortium of microorganisms, bacteria, protozoa, and fungi that live com–mensally or symbiotically with a host. Human individuals may have microbiome characteristics that are unique to their individual biology[2-4].

PHARMACOMETABONOMICS. The prediction of the quantitative outcome of a healthcare (typically drug) intervention in a given individual based on a pre–dose mathematical model of the metabolic state[3-4, 11, 13].

XENOBIOTIC. A compound that is foreign to the endogenous process and has no intrinsic biological function but which can have major effects on endogenous pathway control and can be extensively metabolized by complexes of host enzymic systems that have collectively relatively low substrate specificities[5].

XENOMETABOLOME. Characteristic profile of non–endogenous compounds (drugs and their metabolites, pollutants, dietary components, herbal medicines) in a biofluid[3-4].

REFERENCES.

1. Issaq, H. J.; Abbott, E.; Veenstra, T. D., Utility of separation science in metabolomic studies. *J Sep Sci* **2008,***31* (11), 1936–47.

2. Nicholson, J. K.; Holmes, E.; Wilson, I. D., Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* **2005,***3* (5), 431–8.

3. Nicholson, J. K., Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* **2006,***2*, 52.

4. Holmes, E.; Loo, R. L.; Stamler, J.; Bictash, M.; Yap, I. K.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I. J.; Veselkov, K. A.; Daviglus, M. L.; Kesteloot, H.; Ueshima, H.; Zhao, L.; Nicholson, J. K.; Elliott, P., Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **2008,***453* (7193), 396–400.

5. Nicholson, J. K.; Wilson, I. D., Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat Rev Drug Discov* **2003,***2* (8), 668–76.

6. Han, X.; Gross, R. W., Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry: a bridge to lipidomics. *J Lipid Res* **2003,***44* (6), 1071–9.

7. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry–based metabolomics. *Mass Spectrom Rev* **2007,***26* (1), 51–78.

8. Koek, M. M.; Jellema, R. H.; van der Greef, J.; Tas, A. C.; Hankemeier, T., Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. *Metabolomics* **2011,***7* (3), 307–328.

9. Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D., The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* **2004,***22* (10), 1268–74.

10. Nicholson, J. K.; Lindon, J. C.; Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999,***29* (11), 1181–9.

11. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* **2002,***1* (2), 153–61.

12. Jia, W.; Li, H.; Zhao, L.; Nicholson, J. K., Gut microbiota: a potential new territory for drug targeting. *Nat Rev Drug Discov* **2008,***7* (2), 123–9.

13. Gavaghan, C. L.; Holmes, E.; Lenz, E.; Wilson, I. D.; Nicholson, J. K., An NMR–based metabonomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL10J and Alpk:ApfCD mouse. *FEBS Lett* **2000,***484* (3), 169–74.

14. Clayton, T. A.; Lindon, J. C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J. P.; Le Net, J. L.; Baker, D.; Walley, R. J.; Everett, J. R.; Nicholson, J. K., Pharmaco–metabonomic phenotyping and personalized drug treatment. *Nature* **2006,***440* (7087), 1073–7.

ABBREVIATIONS.

1D	one dimension
^{1}H NMR	proton nuclear magnetic resonance
2D	two–dimensional
3–MPA	3–mercaptopropionic acid
ACN	acetonitrile
AMDIS	Automated Mass Spectrometry Deconvolution and Identification System
ANOVA	analysis of variance test
ANP	aqueous normal phase
AP	acute pancreatitis
APCI	atmospheric pressure chemical ionization
APDS	N,N′–dihydroxysuccinimidyl carbonate and 3–aminopyridine
API	atmospheric pressure interphace
APPI	atmospheric pressure photo ionization
AQC	6–aminoquinolyl–N–hydroxysuccinimidyl carbamate
BD	bipolar disorder
BIOCYC	Pathway/Genome Databases and Pathway Tools Software
BM	breast milk
BMI	body mass index
C^{4}D	contactless conductivity detection
CBQCA	3–(4–carboxybenzoyl)–2–quinolinecarboxaldehyde
CE	capillary electrophoresis
CE	cholesterol esters
CEC	capillary electrochromatography mode
CE–LIF	capillary electrophoresis – laser induced fluorescence
CE–LIF	capillary electrophoresis with laser induced fluorescence detector
CE–MS	capillary electrophoresis coupled to mass spectrometry

CE−NMR–SHY	capillary electrophoresis−nuclear magnetic resonance statistical hetero–analytical correlation spectroscopy
CER	ceramides
CE–UV	capillary electrophoresis –ultraviolet detector
CGE	capillary gel electrophoresis mode
CI	chemical ionization
CID	collision induced disocciation
CIEF	capillary isoelectric focusing mode
CITP	capillary isotachophoresis
CLP	loadings column plot
COSY	correlation spectroscopy
COW	correlation optimized warping method
CPMG	Carr–Purcell–Meiboom–Gill
Ctr–scaling	centre scaling
CV	coefficient of variation
CV1	internal cross validation
CV2	external validation
CZE	capillary zone electrophoresis
D_2O	deuterium water
DA	discriminant analysis
DG	diacylglycerides
DC	direct current
DIMS	direct infusion to mass spectrometry
DModX	distance to model plot
DNA	deoxyribonucleic acid
EC	electrochemical array
ECD	electron–capture detector
ED	electrochemical detection
EI	electron impact ionization
EIC	extracted ion chromatogram

EOF	electroosmotic flow
ESI	electro spray ionization
FAME	fatty acid methyl esters
FbF	Find by Formula
FDR	false discovery rate
FFA	free fatty acids
FID	free induction decay
FID	flame–ionization detector
FPD	flame photometric detector
FT–ICR	Fourier transform ion cyclotron
FT–IR	Fourier transform infrared spectroscopy
FWER	Bonferroni correction and family wise error rate
GC	gas chromatography
GC × GC	multidimensional gas separation system
GC x GC–MS (2D GC–MS)	2 dimensional gas chromatography coupled to mass spectrometry
GC–MS	gas chromatography coupled to mass spectrometry
GL	glycolipids
HCl	hydrochloric acid
HILIC	hydrophilic interaction chromatography
HMBC	heteronuclear multiple–bond correlation spectroscopy
HMDB	Human Metabolome Data Base
HMQC	heteronuclear multiple–quantum correlation spectroscopy
HR–MAS NMR	high–resolution magic–angle–spinning nuclear magnetic resonance
HSQC	heteronuclear single–quantum correlation spectroscopy
ICA	independent component analysis
ID	identification
IS	internal standard
IT	ion trap
JRES	J–resolved

KEGG	Kyoto Encyclopedia of Genes and Genomes
KET	ketamine
***k*-NN**	*k*–nearest neighbor algorithm
KO	knockout
LC	liquid chromatography
LC X LC	multidimensional liquid separation system
LC x LC–MS (2D LC–MS)	2 dimensional liquid chromatography coupled to mass spectrometry
LC–CE	liquid chromatography – capillary electrophoresis
LC–EC	liquid chromatography with electrochemical array detection
LC–MS	liquid chromatography coupled to mass spectrometry
LC–MS/MS	liquid chromatography – tandem mass spectrometry
LC–NMR	liquid chromatography coupled to nuclear magnetic resonance
LED–IF	laser emitting diode–induced fluorescence
LOD	limit of detection
LOESS	locally weighted scatter plot smoothing
LOQ	limit of quantification
LPA	lysophosphatidic acid
LPC	lysophosphatidylcholine
LPE	lysophosphatidylethanolamine
LPG	lysophosphatidylglycerol
LPI	lysophosphatidylinositol
LPS	lysophosphatidylserine
LVs	latent variables
m/z	mass/charge ratio
MALDI	matrix–assisted laser desorption
MASSTRIX	Mass TRanslator into Pathways
MathDAMP	Mathematica package for Differential Analysis of Metabolite Profiles (free software)
MCF	methyl chloroformate

MDD	major depressive disorder
MDS	multi–dimensional scaling
MECK	micellar electrokinetic chromatography
Metalign	free software program for the pre–processing and comparison of full scan nominal or accurate mass LC–MS and GC–MS data.
METLIN	Scripps Center For Metabolomics and Mass Spectrometry
MFE	Molecular Feature Extractor
***m*–HPPS**	*m*–hydroxyphenylpropionyl sulfate
MHz	megahertz
MPP	Mass Profiler Professional
MRM	multiple reaction monitoring
mRNA	messenger – ribonucleic acid
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSTS	mass spectrometry –total signal
MSTUS	mass spectrometry –total useful signal
MWA	metabolome–wide association
MZmine	free software for mass–spectrometry data processing
NBD–F	4 –fluoro–7–nitro–2,1,3–benzoxadiazole
NDA	naphthalene dicarboxaldehyde
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NMDAR	N–methyl–D–aspartate receptor
NMR	nuclear magnetic resonance
NP	normal phase
NPD	nitrogen/phosphorus detector
NRs	non–responders
OPA	*O*–phthaldialdehyde
OPLS	ortogonal partial least squares
OPLS–DA	ortogonal partial least squares – discriminant analysis

PA	phosphatidic acid
PC	principal component
PC	phospholipid
PC	phosphatidylcholine
PCA	principal components analysis
PE	phosphatidylethanolamine
PG	phosphatidylglycerol
PI	phosphatidylinositol
PLA	phenyllactic acid
PLS	partial least squares
PLS–DA	partial least squares – discriminant analysis
ppm	parts per million
PQN	probabilistic quotient normalization
PS	phosphatidylserine
PUFA	polyunsaturated fatty acid
Q	quadrupole
Q1	first quadrupole
Q^2	prediction power
q2	second quadrupole
QC	quality control
QqQ	triple quadrupole
QTOF	quadrupole time of flight
Q–Trap	quadrupole linear ion trap
R^2	goodness of sample classification
RF	radio frequency
RI	retention index
RNA	ribonucleic acid
RP	reversed phase
Rs	responders
RSPA	recursive segment–wise peak alignment

RT	retention time
RTL	retention time locking
SHY	statistical heterospectroscopy
SM	sphingomyelins
SPE	solid–phase extraction
SRM	selected reaction monitoring
STOCSY	statistical total correlation spectroscopy
SUS	shared and unique structure plot
SVM	support vector machine
TAG	triacylglycerides
TCA	tricarboxylic acid cycle
TCD	thermal–conductivity detector
TIC	total ion chromatogram
tm	migration time
TOCSY	total correlation spectroscopy
TOF	time of flight
TSP	trimethylsilyl–2,2,3,3–tetradeuteropropionic acid
UHPLC	ultra–high performance liquid chromatography
UPLC	ultra–performance liquid chromatography
UPLC–MS	ultra–performance liquid chromatography coupled to mass spectrometry
UV/Vis	ultravioleta / visible
UV–scaling	unit variance scaling
Var	variable
VIP	variable importance plot
VPA	valproic acid
XCMS	free software for mass–spectrometry data processing
δ	chemical shift