UNIVERSIDAD DE ALCALÁ
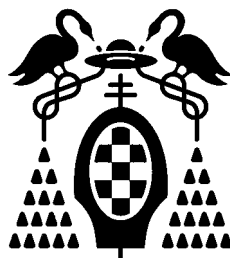
FACULTAD DE BIOLOGÍA, CIENCIAS AMBIENTALES Y QUÍMICA

Departamento de Química Analítica, Química Física e Ingeniería Química

DESARROLLO DE NUEVAS ESTRATEGIAS METABOLÓMICAS Y SU APLICACIÓN EN BIOANÁLISIS Y "FOODÓMICA"

CIAL

CSIC
Consejo Superior de Investigaciones Científicas

CLARA IBÁÑEZ RUIZ

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS DE LA ALIMENTACIÓN
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Julio 2013

CIAL

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

UNIVERSIDAD DE ALCALÁ

FACULTAD DE BIOLOGÍA, CIENCIAS AMBIENTALES Y QUÍMICA

Departamento de Química Analítica, Química Física e Ingeniería Química

DESARROLLO DE NUEVAS ESTRATEGIAS METABOLÓMICAS Y SU APLICACIÓN EN BIOANÁLISIS Y "FOODÓMICA"

Memoria presentada por

CLARA IBÁÑEZ RUIZ

para optar al grado de

DOCTOR INTERNACIONAL

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS DE LA ALIMENTACIÓN

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Trabajo realizado bajo la dirección de:

Alejandro Cifuentes Gallego

Instituto de Investigación en Ciencias de la Alimentación, CSIC

Carolina Simó Ruiz

Instituto de Investigación en Ciencias de la Alimentación, CSIC

CIAL

CSIC
Consejo Superior de Investigaciones Científicas

El Dr. Alejandro Cifuentes Gallego, Profesor de Investigación del Consejo Superior de Investigaciones Científicas (CSIC) en el Departamento de Bioactividad y Análisis de Alimentos del Instituto de Investigación en Ciencias de la Alimentación (CIAL), y la **Dra. Carolina Simó Ruiz**, Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC) en el Departamento de Bioactividad y Análisis de Alimentos del Instituto de Investigación en Ciencias de la Alimentación (CIAL),

CERTIFICAN:

Que el trabajo descrito en la presente memoria, titulado **DESARROLLO DE NUEVAS ESTRATEGIAS METABOLÓMICAS Y SU APLICACIÓN EN BIOANÁLISIS Y "FOODÓMICA",** ha sido realizado bajo su dirección por **Clara Ibáñez Ruiz** en el Laboratorio de Foodómica en el Departamento de Bioactividad y Análisis de Alimentos del Instituto de Investigación en Ciencias de la Alimentación (CIAL) y en el laboratorio de Metabolómica de la Universidad de California (EEUU). Asimismo, autorizan su presentación para que sea defendido como Tesis Doctoral.

Y para que conste y surta los efectos oportunos, firman la presente en Madrid a 10 de Mayo de 2013.

El Dr. D. Jesús Alberto Escarpa Miguel, Profesor Titular de Química Analítica y Director del Departamento de Química Analítica, Química Física e Ingeniería Química de la Universidad de Alcalá

CERTIFICA:

Que el trabajo descrito en la presente memoria, titulado **DESARROLLO DE NUEVAS ESTRATEGIAS METABOLÓMICAS Y SU APLICACIÓN EN BIOANÁLISIS Y "FOODÓMICA",** ha sido realizado por Clara Ibáñez Ruiz en el Laboratorio de Foodómica en el Departamento de Bioactividad y Análisis de Alimentos del Instituto de Investigación en Ciencias de la Alimentación (CIAL) del Consejo Superior de Investigaciones Científicas (CSIC) y en el laboratorio de Metabolómica del Prof. Oliver Fiehn de la Universidad de California (EEUU), bajo la dirección del Prof. Alejandro Cifuentes Gallego y de la Dra. Carolina Simó Ruiz. Asimismo, autoriza su presentación para que sea defendido como Tesis Doctoral.

Y para que conste y surta los efectos oportunos, firma la presente en Alcalá de Henares a 10 de mayo de 2013.

No soy la única, pero aún así soy alguien.

No puedo hacer todo, pero aún así puedo hacer algo;

y justo porque no lo puedo hacer todo, no renunciaré a hacer lo que sí puedo.

Hellen Keller

When a thing was new, people said, 'It is not true'.

Later, when the truth became obvious, people said,

'Anyway, it is not important.' And when its importance

could not be denied, people said, 'Anyway, it is not new'.

William James

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a mis directores de tesis, Carolina y Alejandro, por darme la oportunidad de incorporarme a su grupo de investigación hace cuatro años. Gracias por toda la ayuda y ejemplo que he recibido por vuestra parte, tanto en el aprendizaje de las tareas a nivel científico como en los valores de trabajo duro, constancia y optimismo para poder llegar a ser algún día una buena investigadora. Me he sentido afortunada de contar con vosotros desde el día en que pisé el antiguo Instituto de Fermentaciones Industriales (IFI) y ahora que escribo esta memoria me siento orgullosa del trabajo, que entre los tres, hemos sacado adelante. Gracias por vuestra inmensa paciencia y vuestro gran apoyo.

Me gustaría agradecer al Ministerio de Economía y Competitividad la beca predoctoral FPI que he percibido estos cuatro años y la ayuda a las estancias breves en el extranjero que disfruté en el laboratorio de Oliver Fiehn, a quien también agradezco haberme recibido en su laboratorio.

Me gustaría agradecer a la Prof. Elena Ibáñez por su ayuda y ánimo durante todos estos años al margen de que, a pesar de las habladurías, no seamos familia.

Durante este periodo he coincidido con mucha gente en el laboratorio, tanto gente del extranjero (Joana, Luci, Zeineb, Giovanni, Mustafá) como de la plantilla del laboratorio de Foodómica (Virginia, Alberto, María, Miguel, Jose, Marina, Lidia, Carlos, Pepe, etc) a los que tengo que dar las gracias por haberme hecho mis días difíciles algo más fáciles. Echaré de menos esos cafés, anécdotas, charlas de desestrés y desahogo que me daban aire para seguir. Lástima que como dijo el gran Jose Luis Sampedro "todo lo que no sea alumbrar un futuro basado en la investigación y no en el trabajo precario, es destruir el futuro del país" y dado el poco futuro que nos dejan aquí nos vamos a perder de vista un tiempo. Espero poder volver a trabajar con vosotros cuando sea posible. Gracias por estos cuatro añazos.

A Pedro le tengo que agradecer todo lo que me ha enseñado de estadística, sin desesperarse a pesar de mi ignorancia. Gracias por haber estado siempre disponible

ÍNDICE

LISTA DE ABREVIATURAS

α: nivel de significación

Aβ (*amyloid beta*): β-amiloide

ABN (*acidic, basic and neutral*): ácido, básico y neutro

ACh (*acetylcholine*): acetilcolina

AD (*Alzheimer's disease*): enfermedad de Alzheimer

AEBSF (*4-(2-aminoethyl) benzenesulfonyl fluoride*): 4-(2-aminoetil) bencenosulfonil fluoruro

AICD (*amyloid precursor protein intracellular domain*): dominio intracelular de la proteína precursora amiloidea

ANOVA (*analysis of variance*): análisis de la varianza

APCI (*atmospheric pressure chemical ionization*): ionización química a presión atmosférica

API (*atmospheric pressure ionization*): ionización a presión atmosférica

ApoE (*apolipoprotein E*): apolipoproteína E

APP (*amyloid precursor protein*): proteína precursora amiloidea

β-CD (*β-cyclodextrin*): β-ciclodextrina

CE (*capillary electrophoresis*): electroforesis capilar

CEC (*capillary electrochromatography*): electrocromatografía capilar

CGE (*capillary gel electrophoresis*): electroforesis capilar en gel

CIEF (*capillary isoelectric focusing*): isoelectroenfoque capilar

CITP (*capillary isotachophoresis*): isotacoforesis capilar

CNS (*central nervous system*): sistema nervioso central

CRC (*colorectal cancer*): cáncer colorrectal

CSF (*cerebrospinal fluid*): fluido cefalorraquídeo

CV (*canonical variable*): variable canónica

CVA (*canonical variate analysis*): análisis canónico discriminante

CZE (*capillary zone electrophoresis*): electroforesis capilar zonal

DART (*direct analysis in real time*): análisis directo en tiempo real

DESI (*desorption electrospray ionization*): ionización por desorción con electrospray

DFI (*direct flow injection*): inyección por infusión directa

DMH (*1,2-dimethylhydrazine*): 1,2-dimetilhidrazina

DNA (*deoxyribonucleic acid*): ácido desoxirribonucleico

DSM (*diagnostic and statistical manual of mental disorders*): manual diagnóstico y estadístico de desórdenes mentales

ECA (*electrochemical array detection*): detección electroquímica

EIC (*extracted ion chromatogram*): cromatograma de iones extraídos

EIE (*extracted ion electropherogram*): electroferograma de iones extraídos

ESI (*electrospray ionization*): ionización por electrospray

FIA (*flow injection analysis*): análisis por inyección en flujo

FT-ICR-MS (*Fourier transform ion cyclotron resonance mass spectrometry*): espectrometría de masas de resonancia de ion-ciclotrón con transformada de Fourier

GABA (*γ-aminobutyric acid*): ácido γ-aminobutírico

GC (*gas chromatography*): cromatografía de gases

GCxGC (*comprehensive two-dimensional gas chromatography*): cromatografía de gases bidimensional completa

GSH (*reduced glutathione*): glutatión reducido

GSSG (*oxidized glutathione*): glutatión oxidado

HDAC (*histone deacetylase*): histona desacetilasa

HILIC (*hydrophilic interaction chromatography*): cromatografía de interacción hidrofílica

HMDB (*human metabolome database*): base de datos del metaboloma humano

HPLC (*high-performance liquid chromatography*): cromatografía de líquidos de alta eficacia

HR-MAS (*high resolution magic angle spinning*): espectroscopía de alta resolución de ángulo mágico

ICP (*inductively coupled plasma*): plasma de acoplamiento inductivo

IPA (*Ingenuity pathways analysis*): análisis de rutas por Ingenuity

KEGG (*Kyoto encyclopedia of genes and genomes*): enciclopedia de genes y genomas de Kyoto

LAESI (*laser ablation electrospray ionization*): ablación por láser e ionización por electrospray

LC (*liquid chromatography*): cromatografía de líquidos

LCxLC (*comprehensive two-dimensional liquid chromatography*): cromatografía de líquidos bidimensional completa

LDA (*linear discriminant analysis*): análisis lineal discriminante

LIF (*laser-induced fluorescence*): fluorescencia inducida por láser

LLE (*liquid-liquid extraction*): extracción líquido-líquido

LMPD (*LipidMaps proteome database*): base de datos del proteoma de LipidMaps

LMSD (*LipidMaps structure database*): base de datos de estructuras de LipidMaps

LOO-CV (*leave-one-out cross-validation*): validación cruzada dejando uno fuera

LSD (*least significant difference*): mínima diferencia significativa

LTQ (*linear trap quadrupole*): cuadrupolo de trampa lineal

m/z (*mass-to-charge ratio*): relación masa-carga

MALDI (*matrix-assisted laser desorption/ionization*): desorción/ionización por láser asistida por una matriz

MCI (*mild cognitive impairment*): deterioro cognitivo leve

MCI-S (*stable mild cognitive impairment*): deterioro cognitivo leve y estable

MCI-AD (*mild cognitive impairment with AD progression*): deterioro cognitivo leve con progresión a Alzheimer

MEKC (*micellar electrokinetic chromatography*): cromatografía electrocinética micelar

METLIN (*metabolite mass spectral database*): base de datos de espectros de masas de metabolitos

MH: Mass Hunter software

MPP: Mass Profiler Professional software

MRI (*magnetic resonance imaging*): resonancia magnética de imagen

mRNA (*messenger RNA*): RNA mensajero

MS (*mass spectrometry*): espectrometría de masas

MSS (*microsatellite stability*): estabilidad microsatélite

MS/MS (*tandem mass spectrometry*): espectrometría de masas en tándem

MSI (*mass spectrometry imaging*): espectrometría de masas de imagen

MTA (*methyl-thio-adenosine*): metiltioadenosina

NCBI (*national center for biotechnology information*): centro nacional de información biotecnológica

NFT (*neurofibrillary tangles*): ovillos neurofibrilares

NIMS (*nanostructure-initiator mass spectrometry*): espectrometría de masas con iniciador basado en nanoestructuras

NMR (*nuclear magnetic resonance*): resonancia magnética nuclear

NPV (*negative predictive value*): valor predictivo negativo

PCA (*principal component analysis*): análisis de componentes principales

PET (*positron emission tomography*): tomografía por emisión de positrones

PPV (*positive predictive value*): valor predictivo positivo

Q (*quadrupole*): cuadrupolo

QqQ (*triple quadrupole*): triple cuadrupolo

RNA (*ribonucleic acid*): ácido ribonucleico

RP (*reversed-phase*): fase inversa

SCI (*subjective cognitive impairment*): deterioro cognitivo subjetivo

SIMS (*secondary ion mass spectrometry*): espectrometría de masas de iones secundarios

SPE (*solid phase extraction*): extracción en fase sólida

SPECT (*single-photon emission computed tomography*): tomografía computarizada por emisión de fotones individuales

SSAT (*spermidine/spermine N-acetyltransferase*): espermidina/espermina N-acetiltransferasa

TIC (*total ion chromatogram*): cromatograma de iones totales

TIE (*total ion electropherogram*): electroferograma de iones totales

TOF (*time of flight*): tiempo de vuelo

UPLC (*ultra-high pressure liquid chromatography*): cromatografía de líquidos de ultra-alta presión

ENGLISH SUMMARY

"Omics" technologies refer to a group of advanced analytical technologies used in a high-throughput manner to explore the composition, roles, and relationships of a variety of molecules in a biological system. From a broad point of view, omics technologies include Genomics, Transcriptomics, Proteomics and Metabolomics technologies, which allow the generation of large amounts of data from gene sequence and expression to protein and metabolite patterns, uncovering variability in cellular networks and function of a whole organism (Nicholson and Lindon, 2008; Wilke et al., 2008). Thus, omics technologies significantly improve the simplistic and reductionist experimental models that offer only a temporal snapshot of the huge complexity and dynamic nature of biological networks that govern human health and disease (Ozdemir et al., 2009). The impressive technological development of the high-throughput tools used in these novel omic technologies has brought about a renewed interest in dietary components that may potentially affect gene expression and the integrative physiological and metabolic functions of an organism, linking even more closely the biomedical and food research areas. To meet this major demand, the new discipline of Foodomics has recently emerged (Cifuentes, 2009; Herrero et al., 2010; Herrero et al., 2011) in response to the rising interest to improve human health through nutrition together with a growing demand of more powerful analytical strategies for food safety assessment, traceability and quality control.

This PhD Thesis focuses on the application of Metabolomics in specific biomedical and food research areas. Metabolomics, the newest of the omics technologies, deals with the comprehensive study of the entire set of small (about <1500 Da) molecules (i.e. metabolites) in a biological system (cell, biofluid, tissue, or organism) at a time, under given conditions (Fiehn, 2001). The investigation of the metabolic states of biological systems offers precious information in clinics and dietary intervention studies because metabolites are not merely the end products of gene expression result of the genome-environment interactions, but also form part of the regulatory system in an integrated manner.

The overall aims of this PhD Thesis work are therefore the following ones:

- To develop and optimize advanced analytical methodologies for Metabolomics of cell culture and biological fluids.

- To integrate the results from several metabolomic platforms to widen the coverage of the biological system under study.

- To achieve an optimum data processing, by the development and application of the most appropriate bioinformatic tools according to the analytical platform used and biological sample under study. Simultaneously, to validate and exhaustively inspect the results obtained after every data processing step.

- To design and apply statistical techniques conforming to metabolomic data achieved from data processing.

- To obtain detailed metabolic information on molecular mechanisms variations related to the chemopreventive effects of a rich-polyphenols extract against colon cancer cells, and related to Alzheimer's disease progression.

- To integrate metabolomic results with those obtained from Proteomics and Transcriptomics following a holistic foodomics approach.

Two main research lines can be discerned in this PhD Thesis work accordingly divided in two sections (i.e. **First Section** or "*Primera Parte*" and **Second Section** or "*Segunda Parte*"), containing three and four chapters, respectively.

First Section contains three chapters related to the development, optimization and application of analytical methodologies to study the metabolite content of HT-29 (human colon adenocarcinoma cell line). Antiproliferative effect of a rich-polyphenol extract from rosemary has been investigated following a foodomics approach.

Chapter 1 deals with the comparison of four different metabolite purification procedures from HT-29 cell cultures. Metabolite content from HT-29 cells were obtained after four purification procedures: two solid-phase extractions using a i) conventional C18 sorbent and ii) a polymer-based sorbent, ABN (acidic, basic, neutral), with improved yield on higher polarity molecules extraction; iii) protein precipitation with organic solvent (i.e. methanol); and iv) ultrafiltration via a 3 kDa cut-off membrane filter. After capillary electrophoresis (CE) coupled with time of flight mass spectrometry (TOF MS), detection and comparison of the four metabolic profiles were determined and discussed in terms of metabolites coverage and content.

In **Chapter 2** the antiproliferative effects of a rosemary rich-polyphenol extract

against HT-29 were investigated at a molecular level. Differentially expressed metabolites were identified by three complementary analytical platforms in order to widen the metabolites coverage, namely, CE, reversed-phase ultra-high pressure liquid chromatography (RP/UPLC) and hydrophilic interaction ultra-high pressure liquid chromatography (HILIC/UPLC), all coupled with TOF MS.

A global foodomic study to evaluate the chemopreventive effect of a rosemary extract with antiproliferative activity is presented in **Chapter 3**. Data from Transcriptomics, Proteomics and Metabolomics technologies were examined to obtain a non-target overview of the main molecular processes involved in the reduced proliferation of HT-29 colon cancer cells after the treatment with a rich-polyphenol extract.

The **Second Section** contains four chapters linked to the search of key metabolic biomarkers involved in the progression of Alzheimer's disease (AD).

The potential of Metabolomics in the discovery of biomarkers related to AD progression and diagnosis is reviewed in **Chapter 4** including description and discussion of present trends in AD investigation as well as specific needs and future directions in this research field.

Metabolic differences in cerebrospinal fluid (CSF) samples from individuals in four different disease stages were investigated in **Chapters 5, 6** and **7**. Samples were obtained from subjects who received an initial diagnosis of subjective cognitive impairment (SCI or control), mild cognitive impairment (MCI) and Alzheimer's disease (AD). After a follow-up period of 2 years, a subset of MCI individuals developed AD (MCI-AD) and the rest of the patients remained stable after the initial diagnosis (MDI-S).

Chapter 5 is focused on the optimization of an enantioselective analytical procedure based on chiral micellar electrokinetic chromatography with laser-induced fluorescence detection (chiral-MEKC-LIF) method to determine the differences in D- and L-amino acid content among the four different AD states following a metabolic profiling approach.

Chapter 6 and **Chapter 7** deal with the determination of new early biomarkers of AD progression by CE-TOF MS (**Chapter 6**), and by RP/UPLC- and HILIC/UPLC-TOF MS (**Chapter 7**), following a non-targeted metabolomic approach. Multivariate statistical techniques were applied in both chapters to create predictive models uncovering the potential metabolic biomarkers involved in AD progression.

At the end of the Second Section, a **General Conclusions and Future Work** part is included, consisting of the most relevant concluding remarks and future research lines arising from the publications presented in this PhD work.

Finally, **Appendix I** comprises a list of all the scientific publications that have been generated from the work done during the period of this PhD Thesis.

REFERENCES

Cifuentes A.; *J Chromatogr A* 1216 (2009) 7109.

Fiehn O.; *Comp Funct Genomics* 2 (2001) 155-168.

Herrero M., García-Cañas V., Simó C., Cifuentes A.; *Electrophoresis* 31 (2010) 205-228.

Herrero M., Simó C., García-Cañas V., Ibáñez E., Cifuentes A.; *Mass Spectrom Rev* 31 (2011) 49-69.

Nicholson J.K., Lindon J.C.; *Nature* 455 (2008) 1054-1056.

Ozdemir V., Suarez-Kurtz G., Stenne R., Somogyi A.A., Kayaalp S.O., Kolker E.; *OMICS* 13 (2009) 43-62.

Wilke R.A., Mareedu R.K., Moore J.H.; *Curr Pharmacogenomics Person Med* 6 (2008) 150-159.

1. INTRODUCCIÓN

1.1.   LAS TECNOLOGÍAS ÓMICAS

Las tecnologías "ómicas" son un conjunto de tecnologías analíticas de alto rendimiento empleadas en diferentes disciplinas. Entre ellas, la Genómica, la Transcriptómica, la Proteómica y la Metabolómica, permiten detectar a gran escala los genes, genes transcritos, proteínas y metabolitos, respectivamente, en cualquier sistema biológico. Estas nuevas tecnologías permiten generar y agrupar esta información con el fin de obtener una visión lo más completa posible del sistema biológico en estudio.

La Genómica se encarga del estudio sistemático del conjunto de genes (genoma) de un organismo. Dicho estudio comprende desde el análisis, la identificación y la cuantificación del genoma, hasta la investigación de las interacciones que se producen entre los genes de un organismo, o bien entre los genes con el medio ambiente. Las marcas epigenéticas, procesamiento del mRNA, RNA no codificante, las interacciones entre proteínas y las modificaciones post-traduccionales son algunos ejemplos de recientes hallazgos que han puesto de manifiesto que el genotipo y fenotipo de un organismo no están exclusivamente gobernados por la información contenida en el DNA (Altelaar y col., 2013). Por ello, para aumentar nuestro conocimiento de los mecanismos celulares es necesario entender el papel que juegan las proteínas y sus moléculas precursoras (mRNA) en los procesos biológicos. La Transcriptómica se centra en el estudio de las moléculas intermediarias (mRNA) que constituyen los productos de la transcripción del DNA, a partir de los cuales se sintetizan las las proteínas. Por su parte, la Proteómica se encarga del estudio global de las proteínas y péptidos codificados por un organismo (proteoma), así como los efectos de su modificación, interacción y actividad en un sistema biológico. Aunque a partir del transcriptoma se sintetizan las proteínas, el aumento de moléculas de mRNA no siempre se correlaciona con el aumento en los niveles de proteína sintetizada (Gygi y col., 1999). Además, aunque una proteína complete con éxito su síntesis no implica que dicha molécula sea activa enzimáticamente. Por estos motivos, las modificaciones en el transcriptoma o proteoma no siempre se correlacionan con alteraciones bioquímicas a nivel fenotípico (Sumner y col., 2003). La tecnología ómica de más reciente aparición es la Metabolómica, capaz de revelar la respuesta de los sistemas biológicos a la influencia genética, nutricional y ambiental a través del análisis del conjunto de

metabolitos (metaboloma). Dado que el metaboloma es el producto final de la expresión del genoma, permite investigar la relación entre el genotipo y el fenotipo resultante. Al ser la Metabolómica la tecnología principal desarrollada en esta Tesis, será abordada en profundidad en el siguiente apartado de esta memoria (apartado 1.2).

Los avances en las técnicas analíticas, han sido cruciales para el desarrollo de las tecnologías ómicas. Del mismo modo, el desarrollo de herramientas bioinformáticas ha sido una pieza clave tanto para el avance de dichas técnicas como para la interpretación e integración de los resultados obtenidos (Zhang y col., 2008). Todo ello, junto con la mejora de las bases de datos *on-line*, que recogen la información obtenida de los experimentos realizados mediante el empleo de estas tecnologías, han sido aspectos muy importantes para el progreso de las mismas (Song y col., 2008). En paralelo, han surgido otras tecnologías ómicas fruto del avance en el área de las ciencias de la alimentación. Así, la Nutrigenómica estudia la expresión de los genes en relación con la nutrición y el desarrollo de enfermedades asociadas a dicha expresión. Ofrece, por tanto, las claves para un avance eficaz en la obtención de la denominada dieta personalizada, cuyo objetivo radica fundamentalmente en la prevención o tratamiento de enfermedades asociadas a la interrelación entre genes y nutrición. En relación con lo expuesto anteriormente, aparece la Foodómica (también conocida como Alimentómica), una nueva disciplina definida recientemente (Cifuentes, 2009; Herrero y col., 2010; Herrero y col., 2011), la cual, haciendo uso de las técnicas ómicas (principalmente Transcriptómica, Proteómica y Metabolómica), investiga la relación existente entre los alimentos y la salud, y permite abordar de una forma global aspectos relacionados con la calidad, seguridad y trazabilidad de los alimentos. La Foodómica o Alimentómica ofrece un enfoque holístico a través de la investigación de alimentos y/o efectos a nivel molecular asociados a los mismos.

1.2. LA METABOLÓMICA. OPORTUNIDADES Y DIFICULTADES

La Metabolómica es la más reciente de las tecnologías ómicas y se encarga del estudio sistemático del metaboloma o conjunto del metabolitos de un determinado sistema biológico. Los metabolitos pueden definirse como los compuestos de bajo peso

4

molecular (normalmente <1500 Da) presentes en células, tejidos o fluidos biológicos, que son productos e intermediarios de los procesos químicos o enzimáticos resultado del metabolismo celular. Existe cierta controversia en el uso, a menudo indiferente, de los términos "Metabolómica" y "Metabonómica". El término Metabonómica surgió en 1999 (Nicholson y col., 1999) a partir de la raíz griega "meta" (cambio) y "nomos" (leyes o reglas) en relación a la determinación de cambios en el metabolismo a lo largo del tiempo mediante el uso de herramientas quimiométricas. Sin embargo el término "Metabolómica" apareció como una extensión de la palabra metaboloma haciendo referencia al análisis del mismo en ciertas condiciones (Fiehn, 2001), de forma análoga a otras disciplinas como la "Proteómica" y la "Genómica". El término Metabonómica se suele utilizar como sinónimo a la Metabolómica cuando se pretenden obtener perfiles metabólicos, ya que ambos son conceptos cuyo objetivo implica la determinación de los metabolitos en células, fluidos u organismos, siguiendo una estrategia exploratoria no dirigida. Ante dichas semejanzas y diferencias, algunos investigadores consideran que son términos intercambiables (Robertson, 2005), mientras que otros opinan que la diferencia entre ambos es demasiado significativa para hacerlo (Fiehn, 2002). En esta Tesis los dos términos se han empleado ciñéndose estrictamente a las definiciones y a sus diferentes matices propuestos por los autores de cada concepto (ver Tabla 1.1).

Tabla 1.1. Terminología más ampliamente utilizada en Metabolómica.

Término	Definición	Referencia
Metaboloma	Conjunto de moléculas de bajo peso molecular (metabolitos) que se encuentran en el interior o son secretadas por una célula o tejido.	Nicholson y col., 2003
Metabolómica	Análisis completo del metaboloma de un sistema biológico que se encuentra en unas condiciones determinadas.	Fiehn, 2001
Metabonómica	Medida cuantitativa multiparamétrica de las respuestas metabólicas de los sistemas vivos multicelulares a lo largo del tiempo, tras la aplicación de estímulos fisiopatológicos o modificaciones genéticas.	Nicholson y col., 1999
Huella metabólica	Medida global de los metabolitos presentes en un sistema biológico ("*metabolic fingerprint*") o consumidos o secretados por el mismo (exometaboloma o "*metabolic footprint*") con el objetivo de detectar diferencias entre muestras (p. ej. muestras de sujetos sanos y enfermos).	Fiehn, 2002; Kell y col., 2005
Perfil metabólico	Identificación y cuantificación de un número limitado predeterminado de metabolitos relacionados entre sí (por presentar similitudes físico-químicas o por participar en una ruta metabólica concreta).	Fiehn, 2001

El análisis metabolómico permite estudiar el comportamiento dinámico de los sistemas biológicos mediante la creación de redes interconectadas (rutas metabólicas) entre moléculas relacionadas (Syggelou y col., 2012). Se ha demostrado además que la Metabolómica presenta la posibilidad de detectar cambios mínimos en las rutas metabólicas y la alteración de la homeostasis incluso antes de que sea posible detectar ningún cambio a nivel fenotípico (Nambiar y col., 2010) puesto que los metabolitos no son simplemente los productos enzimáticos de las reacciones bioquímicas sino que, de manera integrada, forman parte de la regulación de los procesos bioquímicos que tienen lugar en los sistemas biológicos (Putri y col., 2013). Dado el potencial de esta tecnología el número de investigaciones que emplean una aproximación metabolómica ha aumentado exponencialmente en la última década tal y como se puede observar en la Figura 1.1.
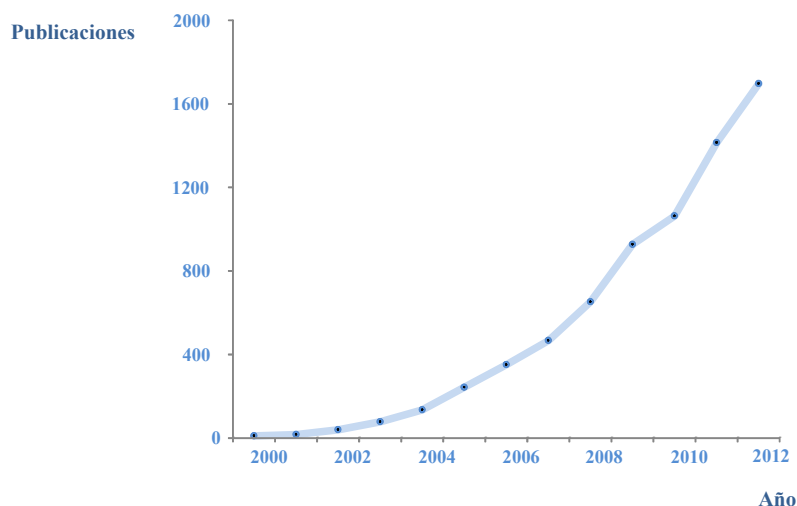
Figura 1.1. Representación gráfica del número de artículos relacionados con la Metabolómica publicados anualmente en el periodo 2000-2012. Búsqueda realizada a través de la base de datos ISI Web of Knowledge (http://apps.webofknowledge.com/) con las palabras claves "Metabolomics" y "Metabonomics".

Fruto del desarrollo de las herramientas ómicas, en la actualidad estamos observando una evolución progresiva que ha permitido ir desde el estudio de los aspectos más puramente clínicos que identifican las diferencias entre individuos sanos o enfermos, hacia una investigación biomédica mucho más integrada que pretende profundizar en los mecanismos moleculares por los que se previene, desencadena, progresa o se trata una determinada patología. La Metabolómica posee un gran potencial en estudios clínicos ya que permite asociar perfiles metabólicos a distintas situaciones fisiológicas y fisiopatológicas. Se considera que está jugando un papel esencial en el conocimiento y diagnóstico de enfermedades (Kussman y col., 2006), ya ha demostrado una gran eficacia en la detección de afecciones respiratorias (Adamko y col., 2012; Snowden y col., 2012), diabetes (Friedrich, 2012), enfermedades cardiovasculares (Brindle y col., 2002; Senn y col., 2012; Shah y col., 2012) y cáncer (Bu y col., 2012; Claudino y col., 2012; Hassanein y col., 2012), entre otras muchas patologías. También está demostrando un gran potencial en los programas de desarrollo de fármacos (Fan y col., 2012). Así por ejemplo, mediante la comparación de huellas/perfiles metabólicos de un estado fisiológico normal (o sano) con otro indicativo de una enfermedad, se pueden detectar las rutas metabólicas alteradas, identificando de

esta manera nuevas dianas terapéuticas, con el consiguiente impacto en el desarrollo de nuevos fármacos. Además la información obtenida mediante los estudios metabolómicos, junto con la información proporcionada por el resto de tecnologías ómicas, son la base para el desarrollo de la denominada medicina personalizada (Eckhart y col., 2012).

Dentro de los estudios preventivos de patologías, o relacionados con la promoción de la salud, está aumentando el interés en la identificación de biomarcadores relacionados con el estilo de vida, con especial énfasis en la investigación nutricional (Whitfield y col., 2004). Uno de los grandes retos actuales que afronta la comunidad científica es la dificultad en la determinación de los efectos de los componentes de la dieta a nivel molecular. La Metabolómica presenta en este sentido una gran oportunidad en la investigación de las relaciones complejas entre la nutrición y el metabolismo (Whitfield y col., 2004) y ha demostrado su potencial en la evaluación de intervenciones dietéticas como se refleja en numerosos trabajos de revisión bibliográfica publicados recientemente (Collino y col., 2011; McNiven y col., 2011; Jones y col., 2012; Llorach y col., 2012). Sin embargo, la influencia de factores tales como posibles sinergias entre ingredientes alimentarios, la gran variabilidad interindividual o la multitud de factores ambientales, dificulta enormemente la interpretación de este tipo de estudios.

Una de las principales dificultades del desarrollo de métodos analíticos siguiendo una estrategia metabolómica no dirigida radica en la naturaleza heterogénea de los metabolitos de cualquier sistema biológico (iones inorgánicos, aminoácidos, ácidos orgánicos, lípidos, etc.), los cuales presentan un amplio intervalo de polaridades, tamaños moleculares, carga a pH fisiológico, etc. Otro factor restrictivo es la acusada diferencia en sus concentraciones, que se estima que puede variar entre 7-9 órdenes de magnitud (pmol-mmol) (Dunn y Ellis, 2005). Dada esta heterogeneidad, no existe un método de preparación de muestra ni una técnica analítica universal que permita la detección de todos los metabolitos presentes en una determinada muestra biológica. Los métodos de preparación de muestra y las herramientas analíticas más empleadas en Metabolómica, así como algunas tendencias interesantes en estos campos, se describirán con más detalle en los apartados 1.2.1 y 1.2.2, respectivamente. Por otro lado, dada la gran resolución y sensibilidad de la instrumentación actual empleada en los estudios metabolómicos, se pueden llegar a generar miles de señales analíticas para cada

muestra, con la dificultad que conlleva su manejo e interpretación (Goodacre y col., 2004). Los análisis metabolómicos de muestras biológicas se traducen en una enorme cantidad de datos que requieren mucho tiempo de análisis computacional (Issaq y col., 2009). Por este motivo el procesamiento y análisis de los datos obtenidos en Metabolómica es un reto que requiere el uso de herramientas bioinformáticas específicas (Mehrotra y Mendes, 2006) y que serán descritas en el apartado 1.2.3. No obstante, no hay que olvidar que para obtener una información válida y útil de los datos ya procesados se hace necesaria la aplicación posterior de herramientas estadísticas adecuadas. El uso del análisis estadístico multivariante está muy extendido en Metabolómica, principalmente aplicado a la selección de los metabolitos más importantes de entre todos los detectados, que permiten diferenciar distintos grupos de muestras, y que son, por lo tanto, potenciales biomarcadores. En el apartado 1.2.4 se describen algunas de las técnicas estadísticas más empleadas en Metabolómica incluyendo las metodologías que se han aplicado en la presente Tesis doctoral. Otra dificultad reside en la identificación de los metabolitos a partir de las señales generadas por las diferentes técnicas analíticas. La identificación se suele llevar a cabo mediante comparación de los datos experimentales con los datos teóricos publicados en bases de datos especializadas. En el apartado 1.2.5 se detalla la metodología y las bases de datos más comúnmente empleadas en los análisis metabolómicos basados en espectrometría de masas.

Teniendo en cuenta estas consideraciones, en toda investigación con un enfoque metabolómico se debe realizar un minucioso y exhaustivo diseño de cada una de las etapas relacionadas con: la recogida y almacenamiento de las muestras, su tratamiento para la purificación de los metabolitos, la elección de la técnica analítica, el procesamiento de los datos, el análisis estadístico, y por último, con la interpretación biológica de los resultados que puedan extraerse del estudio. En cada una de ellas se han de tomar decisiones que desembocarán en el éxito o fracaso del trabajo. En la Figura 1.2 se presentan de manera esquematizada las principales etapas de un estudio metabolómico, las cuales se describen en los siguientes apartados.
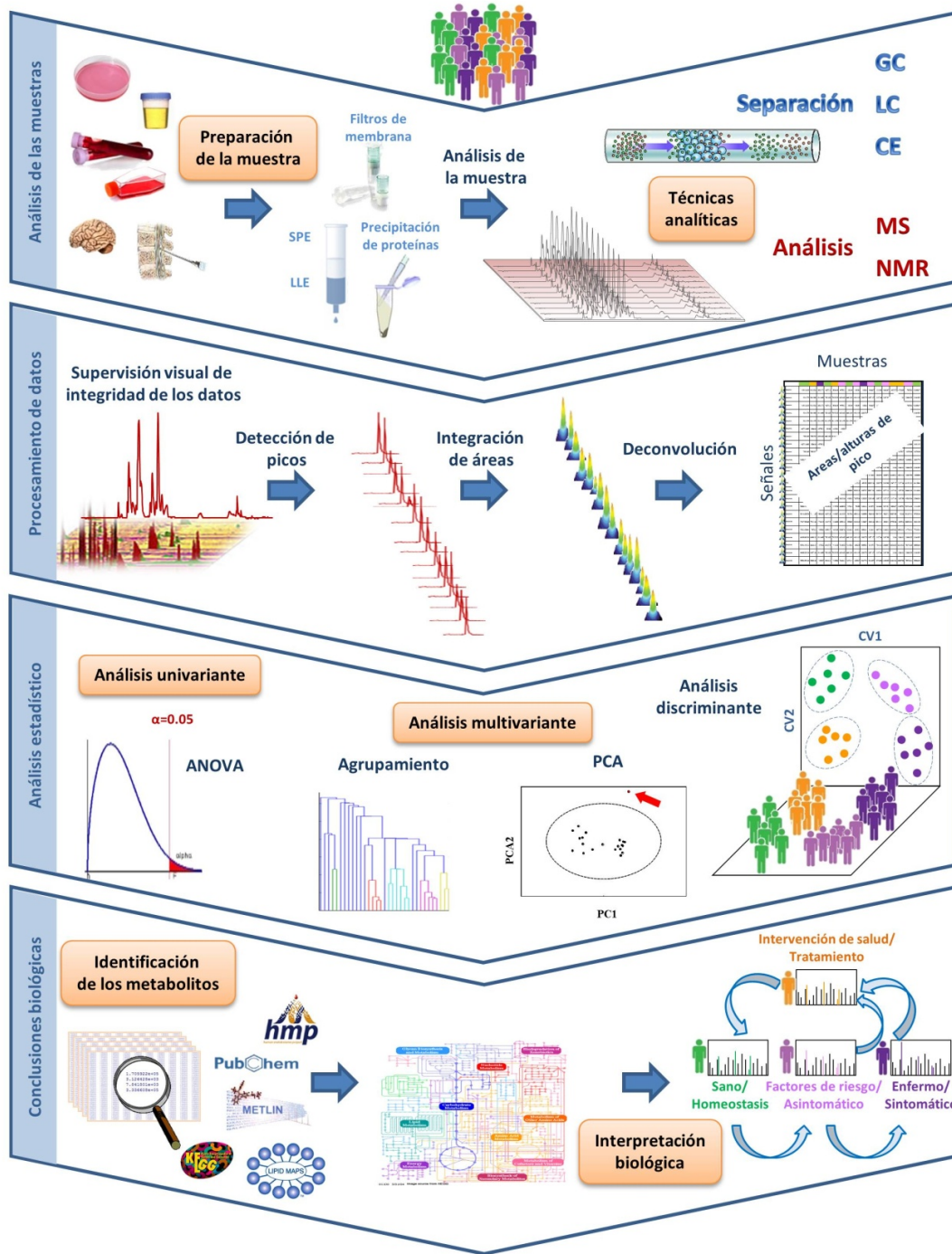
Figura 1.2. Flujo de trabajo genérico de un estudio metabolómico.

1.2.1.   Preparación de la muestra

La elección de la preparación de la muestra es de suma importancia ya que afectará al contenido final de los metabolitos en la muestra purificada, y por consiguiente, a la interpretación biológica de los resultados obtenidos del estudio. El

método ideal de preparación de la muestra en un estudio metabolómico no dirigido debería ser lo menos selectivo, y lo más sencillo y rápido posible para evitar la pérdida o degradación de compuestos durante el tratamiento de la muestra. Además el bloqueo de la actividad enzimática o metabolismo (*"metabolism quenching"*) es fundamental para que la muestra biológica refleje la cantidad real de metabolitos en el momento del muestreo.

El tipo de tratamiento de la muestra previo al análisis metabolómico dependerá fundamentalmente del tipo de muestra biológica y de la plataforma analítica elegida para su análisis. Entre los fluidos biológicos humanos más comúnmente empleados en un estudio metabolómico se encuentran la sangre (plasma y suero), la orina y el líquido cefalorraquídeo (CSF). Además son aquellos en los que se ha detectado un mayor número de metabolitos tal y como se refleja en la base de datos HMDB (*"human metabolome database"*) (Wishart y col., 2009). Por lo general, el tratamiento previo de la muestra está dirigido hacia la eliminación de macromoléculas (DNA y proteínas, fundamentalmente) para evitar interferencias en el análisis químico y/o supresión de las señales analíticas de los metabolitos. En el caso de la orina, que presenta un contenido mínimo de proteínas y una cantidad de compuestos de alto peso molecular relativamente pequeña (Kemperman y col., 2007), el procedimiento más extendido de tratamiento de muestra suele consistir en una centrifugación para eliminar las partículas en suspensión seguida de una dilución de la muestra en función del tipo de técnica analítica (Want y col., 2010), aunque en ocasiones, se emplea un paso previo de filtración (Wong y col., 2008). También se puede añadir azida sódica para evitar el crecimiento de bacterias contenidas en la orina, ya que podrían producir sus propios metabolitos dando lugar a alteraciones en los perfiles metabólicos de la muestra (Ganti y Weiss, 2011). Por último, la urea es un metabolito muy abundante en la orina que puede dificultar la detección o identificación de otros metabolitos con los que es posible que co-migre o co-eluya (Ryan y col., 2011) por lo que en algunos casos se contempla la adición de ureasa en el tratamiento de la muestra. Sin embargo existe bastante controversia en la comunidad científica en cuanto al beneficio de la adición de esta enzima ya que el tratamiento con ureasa interfiere en la detección de ciertos metabolitos (Kind y col., 2007). Con respecto a otros fluidos como el plasma, suero o CSF, la principal dificultad que presentan es su alto contenido proteico. Por este motivo los procedimientos más extendidos para la preparación de las muestras son la precipitación de proteínas mediante la adición de

disolventes orgánicos (acetonitrilo, metanol, etanol, etc.) (Wishart y col., 2008; Bruce y col., 2009) y la ultrafiltración (Tiziani y col., 2008; Rosenling y col., 2011). La adición de disolventes orgánicos no solo provoca la precipitación de las proteínas, sino que también rompe los posibles enlaces entre las proteínas y los metabolitos asociados a las mismas, hecho que no se consigue al aplicar la ultrafiltración como método de preparación de la muestra (Vuckovic, 2012). La precipitación de proteínas con disolventes orgánicos ha sido evaluada en numerosas ocasiones sin llegar a un consenso en cuanto la composición del disolvente orgánico, ni a la proporción del disolvente que se debe añadir a la muestra (Polson y col., 2003; Jiye at al., 2005; Fiehn y col., 2007; Bruce y col., 2009; Pereira y col., 2010). Por otro lado, la ultrafiltración se considera un método de preparación de muestra muy apropiado para la realización de análisis metabolómicos mediante resonancia magnética nuclear (NMR) (Tiziani y col., 2008) y el que proporciona mayor estabilidad en los niveles de ciertos metabolitos en análisis mediante electroforesis capilar acoplada a espectrometría de masas (CE-MS) (Lee y Britz-McKibbin, 2009). Por su parte, la extracción en fase sólida (SPE) es especialmente útil en análisis de perfiles metabólicos ya que se considera un procedimiento más selectivo que los métodos mencionados anteriormente (Michopoulos y col., 2009), aunque también se ha utilizado en estudios metabolómicos no dirigidos (Rezzi y col., 2008). Supone además un método de preparación de muestra especialmente útil cuando se desea concentrar la fracción metabólica de la misma. Por último, para llevar a cabo la extracción de metabolitos en tejidos y cultivos celulares se utilizan distintos disolventes, normalmente tras la congelación y pulverización (en tejidos) de la muestra biológica. El disolvente para la extracción se puede seleccionar según se desee obtener en el extracto metabólico, un predominio de metabolitos polares (isopropanol, etanol, agua, etc.) (Roessner y col., 2000; Aharoni y col., 2002), apolares (cloroformo, etilacetato, etc.) o bien utilizando ambos tipos de disolventes para crear un sistema bifásico y así analizar posteriormente por separado ambos extractos metabólicos (Colebatch y col., 2004; Gullberg y col., 2004). Entre otras muestras biológicas analizadas en estudios metabolómicos se encuentran las heces (Jansson y col., 2009), la saliva (Zhang y col., 2012a), las lágrimas (Chen y col., 2011), el líquido amniótico (Amorini y col., 2012), o la bilis (Wang y col., 2007).

1.2.2.   Estrategias y técnicas analíticas

El análisis del **perfil metabólico** y la determinación de la **huella metabólica** son las dos estrategias más ampliamente aplicadas en Metabolómica. El análisis del perfil metabólico está dirigido al estudio de un grupo de metabolitos con propiedades físico-químicas similares y/o que participan en una determinada ruta metabólica. Este tipo de enfoque es adecuado cuando los metabolitos de interés están predefinidos al inicio del estudio, de modo que todas las etapas de experimentación y análisis de los datos se optimizan para la determinación de dichos compuestos. Por otro lado la obtención de la huella metabólica tiene como objetivo la obtención del mayor número de señales analíticas metabólicas de un determinado sistema biológico. Se trata de la estrategia adecuada si el objetivo del estudio es la clasificación de muestras y/o la búsqueda de biomarcadores. Este tipo de enfoque presenta una dificultad considerablemente mayor al análisis del perfil metabólico ya que para la obtención de la huella metabólica, el diseño de cada etapa del estudio debe permitir el análisis de la mayor cantidad de metabolitos posible, o dicho de otro modo, tiene que evitar en la medida de lo posible la exclusión de la detección de cualquier tipo de metabolito. En la actualidad este hecho es, desde un punto de vista práctico, imposible de alcanzar dado que en cada etapa de un estudio metabolómico se produce un sesgo hacia la detección de unos metabolitos en detrimento de la detección de otros.

Además, no existe una plataforma analítica universal que permita el estudio de todos los metabolitos contenidos en una determinada muestra biológica. Aunque los avances técnicos en la instrumentación así como la hibridación de técnicas analíticas están aportando una información metabolómica cada vez más completa, esta situación sigue siendo un gran reto desde el punto de vista analítico. Por este motivo, buena parte de los esfuerzos en Metabolómica están dirigidos tanto hacia el desarrollo instrumental como al metodológico. En la actualidad, de todo el espectro de técnicas tradicionalmente empleadas en el análisis químico, la NMR y la MS son las que están proporcionando una mayor información a nivel metabolómico, y son por lo tanto, la base de prácticamente cualquier estudio de este tipo.

1.2.2.1. NMR

La NMR es una técnica no destructiva que permite el análisis de muestras muy complejas con un mínimo (o ningún) pre-tratamiento previo. Inicialmente, la NMR se estableció como la técnica de referencia en Metabolómica por tratarse de una técnica no destructiva y por proporcionar información tanto estructural como cuantitativa. Por ello, esta técnica analítica ha sido frecuentemente empleada en la obtención de perfiles y huellas metabólicas (Zhang y Powers, 2012). Al ser la NMR una técnica no selectiva, permite elucidar las estructuras moleculares de los metabolitos sin llevar a cabo una selección *a priori* de los mismos. A pesar de las ventajas que presenta, la principal limitación de la técnica NMR es su relativa baja sensibilidad en comparación con la MS lo que hace que en la mayor parte de los estudios por NMR sólo se detecten los metabolitos más abundantes de las muestras.

1.2.2.2. MS

La MS permite el análisis de compuestos en función de su relación masa-carga (m/z), pudiéndose utilizar para el estudio de la abundancia isotópica, determinación de la masa y de la fórmula molecular, así como para la identificación de compuestos mediante su patrón de fragmentación y para la realización de estudios cuantitativos. La selectividad, sensibilidad e información estructural que proporciona esta técnica hacen de su uso una herramienta eficaz en análisis metabolómicos. El empleo de analizadores de masas de alta y ultra-alta resolución - como por ejemplo: tiempo de vuelo (TOF), espectrometro de masas de resonancia de ion-ciclotrón con transformada de Fourier (FT-ICR-MS), Orbitrap®, etc. - son necesarios en Metabolómica debido a la gran complejidad del metaboloma de cualquier muestra biológica, para cuya elucidación se requiere una elevada exactitud en la determinación de los valores de m/z que sólo es proporcionada por los mencionados analizadores de masas de alta resolución. Además, el empleo de espectrómetros de masas en tándem (MS/MS) permite adquirir espectros de fragmentación de masas de alta y ultra-alta resolución (cuadrupolo (Q)-TOF MS, TOF/TOF MM, cuadrupolo de trampa lineal (LTQ)-Orbitrap, etc.) proporcionando información estructural adicional para su aplicación posterior en el proceso de identificación de los metabolitos de interés.

En general, los resultados obtenidos mediante NMR y MS aportan una información complementaria, necesaria a la hora de llevar a cabo un análisis del metaboloma lo más completo posible. El empleo de estas técnicas en Metabolómica se ha evaluado y comparado en diversas revisiones bibliográficas (Dettmer y col., 2007; Lei y col., 2011; Barding y col., 2012; Smolinska y col., 2012). Dada la importancia de la MS en Metabolómica y siendo además ésta la metodología que se ha utilizado en la presente Tesis doctoral, a continuación se llevará a cabo una descripción de las estrategias más empleadas así como de las tendencias y desarrollos de la MS en los últimos años.

1.2.2.2.1.   Análisis directo mediante MS

El análisis directo mediante MS está adquiriendo cada vez más importancia en aplicaciones de "alto rendimiento" (Huang y col., 2011). Esto se debe fundamentalmente a la gran rapidez del análisis que proporciona. Mediante el empleo de una gran variedad de sistemas de introducción de la muestra se pueden obtener huellas metabólicas por MS en unos pocos segundos, lo cual presenta una ventaja considerable cuando el número de muestras a analizar es elevado. Los sistemas de ionización a presión atmosférica (API) son los más empleados cuando la inyección de la muestra se lleva a cabo por infusión directa (DFI). Entre estos sistemas API, la ionización por electrospray (ESI) y la ionización química a presión atmosférica (APCI) son los más utilizados. Este tipo de introducción de la muestra por infusión directa al espectrómetro de masas requiere una purificación previa de la fracción metabólica, que debe estar disuelta en un disolvente adecuado.

Gracias al desarrollo en los últimos años de analizadores de masas con notables avances en la resolución de las masas, el empleo de MALDI (desorción/ionización por láser asistida por una matriz) para el análisis de moléculas de bajo peso molecular está cobrando cada vez más importancia (Cohen y Gusev, 2002; Van Kampen y col., 2011). Respecto a otros sistemas de introducción de muestra, la principal ventaja del uso de MALDI es una mayor tolerancia a las sales presentes en la misma. Por otro lado, una de las principales limitaciones es la supresión de las señales de los metabolitos debida a los iones que produce la matriz química añadida a la muestra, necesaria para el proceso de desorción/ionización por láser. Por este motivo se están desarrollando diversas

estrategias analíticas en las que se está sustituyendo el uso de las matrices normalmente utilizadas en MALDI por otras soluciones más compatibles con el análisis de compuestos de bajo peso molecular, como por ejemplo, el empleo de matrices basadas en iones líquidos (Vaidyanathan y col., 2006) o en compuestos como la 9-aminoacridina (Edwards y Kennedy, 2005; Miura y col., 2010). Las superficies de silicona porosa se han empleado también con éxito en el análisis de mezclas complejas de metabolitos (Vaidyanathan y col., 2005; Vaidyanathan y col., 2007).

A partir del año 2000, los mayores esfuerzos en el análisis directo mediante MS se han dirigido hacia el desarrollo de las denominadas técnicas de ionización en condiciones "ambiente". Este grupo de técnicas abarcan técnicas de ionización que combinan diferentes tipos de sistemas de introducción y de ionización de la muestra (Venter y col., 2008; Huang y col., 2011; Weston, 2010). La característica común a todas ellas es que son técnicas que permiten el análisis directo de muestras con poco o ningún tratamiento previo y además se llevan a cabo "al aire libre", es decir, no se emplean cámaras de nebulización cerradas. Por esto, las técnicas que emplean API y MALDI no se encuentran entre este grupo de técnicas de ionización, ya que en las primeras es necesario llevar a cabo un exhaustivo tratamiento de la muestra (transferencia de los metabolitos desde la muestra biológica al disolvente adecuado), y en la segunda se require normalmente una cámara de ionización con cierto vacío. El desarrollo de este grupo de técnicas de ionización en condiciones ambiente comenzó en 2004 con la introducción del sistema ESI directo, denominado DESI (desorción e ionización por electrospray) (Takats y col., 2004). Desde entonces se han desarrollado una gran variedad de posibilidades basadas diferentes tipos de sistemas de introducción y de ionización de la muestra. En una reciente revisión bibliográfica sobre el tema se describen más de 30 tipos de técnicas de desorción/ionización ambientales empleadas entre los años 2009 y 2010 en MS (Harris y col., 2011). Entre todas ellas, DESI y DART (análisis directo en tiempo real) son las más empleadas. Aunque todavía el mayor número de aplicaciones empleando este tipo de técnicas de ionización están principalmente dirigidas hacia el análisis de grupos concretos de compuestos, DESI-MS y DART-MS han demostrado su utilidad en el análisis metabolómico de muestras biológicas (Chen y col., 2006) y de alimentos (Novotná y col. 2012) llevando a cabo un tratamiento mínimo de la muestra.

El empleo de espectrometría de masas de imagen (MSI) es otra de las tendencias de los últimos años en en análisis directo por MS en aplicaciones metabolómicas (Miura y col., 2012). Mediante esta técnica se obtienen representaciones espaciales de la abundancia de los metabolitos en secciones/superficies de muestras sólidas. Para llevar a cabo este tipo de análisis se emplean sistemas de desorción/ionización como SIMS (MS de iones secundarios), DESI, NIMS (MS con iniciador basado en nanoestructuras), LAESI (ablación por láser e ionización por electrospray) y MALDI, entre otros. Las características de los métodos de ionización más comúnmente empleados en MSI se recogen en algunas revisiones bibliográficas recientes (Lee y col., 2012; Miura y col., 2012). La aplicación de la MSI al análisis de compuestos de bajo peso molecular presenta un gran potencial y se encuentra en continuo crecimiento (Greer y Sturm, 2011; Svatos, 2010). Cabe destacar aquí la gran importancia que tiene en MSI una rigurosa preparación de la muestra con el fin de obtener datos lo más reproducibles y exactos posibles (Goodwin, 2012).

Por último, es necesario subrayar que los problemas de sensibilidad debidos a los fenómenos de supresión iónica y baja eficiencia en la ionización son frecuentes en el análisis directo por MS. Con el fin de disminuir la complejidad espectral y adquirir la mayor información metabólica posible, la espectrometría de masas suele ir acoplada a técnicas se separación previas, como cromatografía de gases (GC), cromatografía de líquidos (LC) y electroforesis capilar (CE).

1.2.2.2.2.   Acoplamientos GC-MS, LC-MS y CE-MS

El acoplamiento GC-MS se emplea ampliamente en Metabolómica debido a su gran capacidad de separación, robustez instrumental y repetibilidad de los espectros de masas (Zhang y col., 2012b). Una de las consecuencias de esta robustez y repetibilidad es la creación de librerías de tiempos de retención y espectros de masas (obtenidos normalmente empleando la interfase de impacto electrónico entre el cromatógrafo de gases y el espectrómetro de masas) que en la actualidad se manejan y se comparten entre laboratorios. La técnica GC-MS es particularmente apropiada para el análisis de compuestos orgánicos volátiles y otros compuestos no volátiles previamente derivatizados. En la actualidad se emplea de manera muy frecuente en la evaluación de contaminantes en alimentos (McGhie y Rowan, 2012). Sin embargo su potencial va más

allá, y ha demostrado gran utilidad en aplicaciones metabolómicas (Pasikanti y col., 2008; Mitrevski y col., 2009; Yoshida y col., 2012) con un claro predominio de la investigación del metabolismo vegetal sobre los estudios con enfoque biomédico (Fiehn, 2008).

La separación de los compuestos por cromatografía de líquidos (LC) se rige por interacciones químicas entre la muestra, la fase móvil y la fase estacionaria constituida por las partículas de relleno de la columna cromatográfica. Estas interacciones determinan el grado de retención y separación de los componentes contenidos en la muestra. Aunque la resolución de esta técnica es inferior a la que proporciona la cromatografía de gases, permite la separación de compuestos con un mayor peso molecular y polaridad sin la necesidad de una etapa previa de derivatización. En aplicaciones metabolómicas el acoplamiento entre LC y MS se lleva a cabo típicamente mediante ESI, y en menor medida mediante APCI (Theodoris y col., 2010). Dependiendo principalmente del tipo de fase estacionaria elegida, los metabolitos se podrán separar en base a una gran variedad de características físco-químicas (Patti, 2011). En los últimos años se está observando una clara tendencia hacia el empleo de cromatografía de líquidos de ultra-alta presión (UPLC). Los avances que supone el uso de la UPLC frente a la HPLC (cromatografía de líquidos de alta eficacia) se deben fundamentalmente al empleo de partículas, que constituyen la fase estacionaria, con un diámetro interno inferior a 2μm. El resultado es un análisis cromatográfico con una mayor capacidad y resolución (hasta 10^5 platos teóricos/metro), además de proporcionar una mayor sensibilidad en comparación con las columnas de HPLC convencionales reduciendo al mismo tiempo la duración del análisis (Wang y col., 2011). La técnica UPLC-MS se está estableciendo progresivamente como técnica de referencia en Metabolómica, como lo demuestra el rápido aumento en su empleo para el estudio de enfermedades (Roux y col., 2011) y de los efectos de los alimentos/dieta sobre la salud (Puiggròs y col., 2011).

La electroforesis capilar (CE) engloba un conjunto de métodos de separación basados en la diferente movilidad electroforética de los analitos de la muestra bajo la acción de un campo eléctrico en el interior de un capilar, normalmente, de sílice fundida (Ewing y col., 1989). Permite la separación de múltiples moléculas, desde iones inorgánicos hasta macromoléculas, e incluso partículas y células, con eficacias muy

elevadas. Además los volúmenes de muestra y reactivos necesarios para llevar a cabo el análisis son muy pequeños, unos pocos nanolitros y mililitros, respectivamente. El acoplamiento de un detector tan selectivo como la MS a una técnica analítica tan versátil y que proporciona eficacias tan elevadas como la CE da como resultado una potente herramienta de análisis. Existen diversos modos de separación en CE entre los que se encuentran la electroforesis capilar zonal (CZE), la electrocromatografía (CEC), la cromatografía electrocinética micelar (MEKC), la electroforesis capilar en geles o redes poliméricas (CGE), el isoelectroenfoque capilar (CIEF) y la isotacoforesis capilar (CITP). Estas distintas modalidades se diferencian principalmente en la naturaleza del medio de separación que se encuentra dentro del capilar y en las características de los analitos que se pretenden separar. Los modos de CE más adecuados para su acoplamiento con MS son CZE y CEC ya que son los que emplean un tampón de separación más compatible (electrolitos volátiles) con la posterior detección mediante MS. La principal dificultad del acoplamiento CE-MS es el cierre del circuito eléctrico en el interior del capilar de separación. Aunque se han utilizado de forma satisfactoria diversas interfases para el acoplamiento CE-MS (Hommerson y col., 2011), empleando entre otras por ejemplo, la técnica de plasma de acoplamiento inductivo (ICP) (Michalke, 2005) y MALDI (Huck y col., 2005), la interfase ESI en sus distintas modalidades es la más utilizada en la actualidad para llevar a cabo este acoplamiento. Desde el primer acoplamiento CE-MS empleando ESI descrito por Olivares y col. en 1987 (Olivares y col., 1987), en el cual la conexión eléctrica se realizaba mediante un recubrimiento de la punta del capilar de separación con un metal conductor, se han ido desarrollando y perfeccionando diferentes modalidades de ESI para CE-MS (Hommerson y col., 2011). Entre ellas, podemos destacar la interfase ESI con flujo adicional desarrollada inicialmente por el grupo de Smith en 1988 (Smith y col., 1988), por ser la más robusta y la más empleada en la actualidad (Zhao y col., 2012). Tal y como se muestra en la Figura 1.3, esta interfase está formada por tres tubos concéntricos, el primero de ellos, situado en el interior, es el propio capilar de separación que se encuentra rodeado de un tubo de acero inoxidable por el que fluye el flujo adicional. Entre el segundo y tercer tubo fluye un gas inerte, normalmente nitrógeno, que favorece la nebulización de la muestra. El flujo adicional que se hace pasar por el tubo metálico que rodea el capilar cierra el circuito eléctrico necesario durante la separación por CE, al establecer el contacto eléctrico entre el electrodo situado en el vial de entrada, el capilar de separación y el tubo de acero inoxidable (que

funciona como un electrodo).
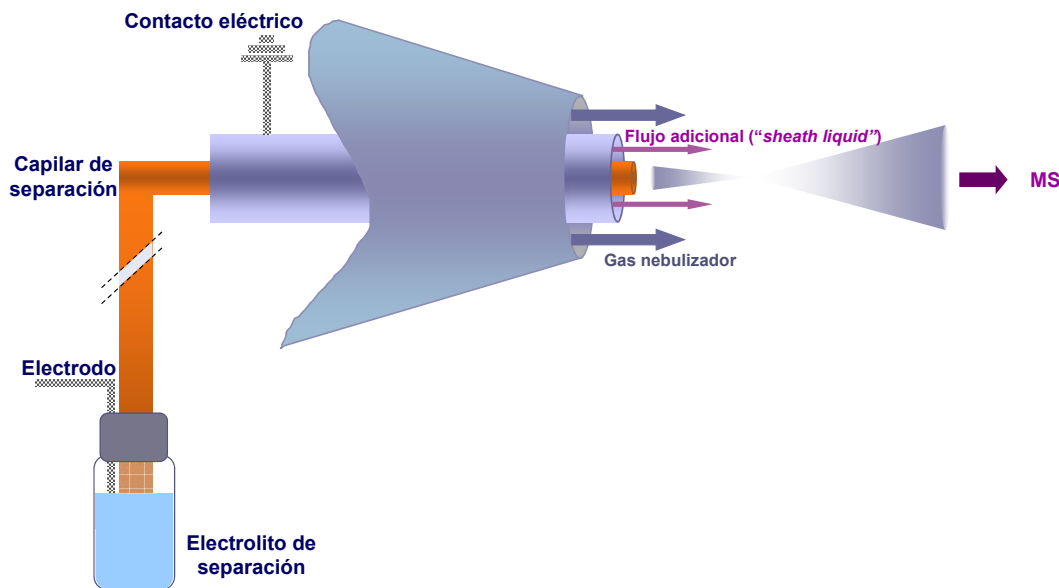
Figura 1.3. Esquema de la interfase electrospray con flujo adicional.

El principal inconveniente de esta interfase es la dilución de los compuestos a analizar debido al empleo del líquido adicional. Esta dilución influye negativamente en la sensibilidad obtenida. Recientemente se ha descrito una interfase sin flujo adicional en la cual el cierre del circuito eléctrico se lleva a cabo mediante el empleo de sílice porosa en el tramo final del capilar de separación (Moini, 2007). De este modo, la última sección del capilar de separación se somete a un tratamiento con ácido fluorhídrico hasta conseguir que la sílice sea porosa. La conexión eléctrica se alcanza insertando la zona porosa del capilar en una aguja metálica (electrodo de salida) rellena de un electrolito de separación. Se evita por lo tanto el empleo de un líquido adicional con la consiguiente mejora en la sensibilidad. Se ha demostrado que el empleo de CE-MS en Metabolómica es una potente herramienta analítica complementaria a LC-MS y GC-MS, ya que permite el análisis de los analitos iónicos o de mayor polaridad que son difíciles de analizar mediante las otras técnicas. En los últimos años se ha publicado un importante número de trabajos mostrando la capacidad de esta técnica para el análisis de perfiles y huellas metabólicas en una gran variedad de muestras (plantas, alimentos, fluidos biológicos, células, tejidos, organismos, etc.) (Ramautar y col., 2009; Ramautar y col., 2011; Ramautar y col., 2013).

La combinación de dos o más separaciones en las denominadas configuraciones multidimensionales es en la actualidad una estrategia analítica cada vez más empleada para mejorar la separación de los metabolitos que co-eluyen en muestras muy complejas. Las separaciones multidimensionales aprovechan el poder de separación de dos o más modos de separación ortogonales, es decir, con una selectividad en la separación complementaria. Estas estrategias multidimensionales, especialmente, en el modo "completo" (por el que todos los compuestos analizados en la primera dimensión se analizan en la segunda), proporciona un aumento en la capacidad de pico, la selectividad y la resolución. La mayor parte de las separaciones multidimensionales completas emplean combinaciones de las técnicas LC y GC. La técnica GC x GC (cromatografía de gases bidimensional completa) es, entre todas las combinaciones, la más empleada. Los principios teóricos y la tecnología empleada se han descrito ampliamente en numerosas revisiones bibliográficas (Pierce y col., 2008; Mostafa y col., 2012; Murray, 2012). La configuración GC x GC más empleada en aplicaciones metabolómicas se basa en el empleo de una columna apolar seguida de una columna polar (Mondello y col., 2008). La combinación LC x LC (cromatografía de líquidos bidimensional completa) está cobrando cada vez más importancia para la separación de muestras complejas de compuestos poco volátiles de bajo peso molecular. En la mayor parte de las aplicaciones publicadas se combinan diferentes modos de separación, siendo las más empleadas la fase reversa, la fase normal, el intercambio iónico y la exclusión molecular (Pol y Hyötyläinen, 2008). En Metabolómica las aplicaciones de la técnica LC x LC han sido hasta la fecha escasas debido principalmente a la necesidad de herramientas más efectivas para el procesamiento de los datos (Kivilompolo y col., 2011) y a que se trata de separaciones que requieren tiempos de análisis muy largos, con lo cual, se obtienen niveles de reproducibilidad menores que con otras técnicas como la GC x GC. Con el fin de reducir el tiempo de análisis total, se han desarrollado algunos acoplamientos novedosos, como por ejemplo, LC-CE-ESI-MS, en el cual la segunda dimensión se lleva a cabo en un chip-CE en unos pocos segundos (Chambers y col., 2011).

1.2.2.3. *Multiplataformas analíticas*

Dada la complejidad del metaboloma de cualquier muestra biológica, el análisis

metabolómico mediante varias plataformas analíticas es una práctica cada vez más frecuente. La combinación de la información procedente de distintas técnicas analíticas se traduce en la ampliación del conocimiento del metaboloma de los organismos en estudio y en el aumento de la probabilidad de éxito en la detección de potenciales biomarcadores entre muestras diferentes. Las multiplataformas han sido ampliamente empleadas en análisis metabolómicos con aplicación clínica (Birungi y col., 2010; Suhre y col., 2010; Saric y col., 2012; Temmerman y col., 2012) y en investigación de los efectos de ingredientes alimentarios sobre la salud (Law y col., 2008; Martin y col., 2009; van Dorsten y col., 2010; Mellert y col., 2011; Jacobs y col., 2012). Un ejemplo de la obtención de información complementaria procedente de LC-MS y CE-MS, y su posterior integración se muestra en el **Capítulo 2** de esta Tesis.

1.2.3.   Procesamiento de los datos

El procesamiento de datos y posterior análisis estadístico en Metabolómica requieren principalmente de la aplicación de dos disciplinas: la quimiometría y la bioinformática. La quimiometría se puede definir como el conjunto de herramientas matemáticas y estadísticas para su aplicación en química (Lavine y Workman, 2004) y se diferencia del término "bioinformática" en que éste último abarca el almacenamiento, obtención, análisis e interpretación de la información generada en un contexto biológico (Bains, 1996). A diferencia del procesamiento de datos en los estudios metabolómicos dirigidos que han sido ampliamente desarrollados, el procesamiento de datos para estudios no dirigidos constituye aún un desafío. En estos estudios se desconoce *a priori* la composición cualitativa y cuantitativa de la muestra a nivel metabólico y por tanto cada una de las etapas del análisis metabolómico ha de ser optimizada, incluyendo el procesado y análisis estadístico de los datos (apartado 1.2.4). Los datos procedentes del espectrómetro de masas se caracterizan por ser altamente complejos y de gran volumen por lo que su procesamiento ha de realizarse con herramientas bioinformáticas especializadas (Mehrotra y Mendes, 2006). En primer lugar, los datos se deben transformar o convertir en formatos compatibles con las herramientas bioinformáticas que se vayan a aplicar posteriormente. Los formatos más comunes son netCDF o mzXML. Una vez que los datos están en el formato adecuado, se procesan con herramientas informáticas específicas. El procesado de datos comprende numerosas

etapas (Sugimoto y col., 2012). De forma general, las etapas principales que se llevan a cabo son: la detección de las especies iónicas (también llamados "*features*"), eliminación de las señales que no proceden de la muestra, alineamiento del tiempo de migración/retención y anotación de los picos detectados. Debido a que los capítulos de esta Tesis contienen datos procedentes principalmente de LC-MS y CE-MS, a continuación se profundiza en el procesamiento de los datos procedentes de dichas técnicas. La gran mayoría de las herramientas informáticas desarrolladas para estudios metabolómicos han sido diseñadas para datos procedentes de LC-MS (Bellew y col., 2006), entre las que se encuentran MZmine (Katajamaa y Oresic, 2005), XCMS (Smith y col., 2006) o CentWave (Tautenhahn y col., 2008). Si bien es cierto que es necesario un mayor desarrollo de dichas herramientas para un correcto procesado de datos de LC-MS, (por ejemplo, disminuir el número de falsos positivos, que por lo general es muy elevado) en el caso de los datos procedentes de CE-MS este procedimiento es mucho más crítico. Las formas de pico obtenidas de CE-MS y su complejidad (en comparación con LC o GC), además de las desviaciones en tiempos de migración entre inyecciones, suponen las dificultades más importantes a las que hacer frente en el procesado de datos procedentes de CE-MS (Sugimoto y col., 2010). La heterogeneidad en la forma de los picos afecta principalmente en la etapa de detección de "*features*". Por este motivo cuando se lleva a cabo una detección completamente automática es necesario realizar una revisión exhaustiva posterior de los resultados obtenidos. Otra opción que resulta de gran ayuda es la aplicación de algoritmos interactivos en los que el usuario pueda ajustar los parámetros y previsualizar los resultados para optimizar esta detección inicial de especies iónicas. En cualquiera de los dos casos, es necesario optimizar cada parámetro aplicado en el algoritmo de detección. Aunque este proceso es más crítico en el caso de datos procedentes de CE-MS, esta revisión y optimización debe realizarse también en el caso de datos procedentes de LC-MS. La detección de señales es la base de un buen procesamiento de datos ya que constituye el primer paso que afectará por lo tanto a todos los pasos sucesivos, y por consiguiente a los resultados finales e interpretación biológica de los mismos. Por su parte, la desviación del tiempo de migración entre inyecciones en CE-MS, debido principalmente a modificaciones en la pared interna del capilar o del electrolito de separación inducido por la matriz de la muestra, se puede considerar el principal inconveniente de CE en comparación con LC o GC. Por este motivo, en el caso de los datos procedentes de CE-MS está especialmente aconsejado el uso de herramientas bioinformáticas versátiles, interactivas

y con un alto número de parámetros a optimizar para conseguir no sólo la detección de "*features*" sino un correcto alineamiento de los tiempos de migración entre las muestras. Entre la oferta, cada vez más amplia, de programas informáticos gratuitos disponibles para estudios metabolómicos, se describen brevemente a continuación los que, por su interés tras una comparación con otras herramientas bioinformáticas, han sido seleccionados para su uso en los capítulos contenidos en esta Tesis.

- MZmine: Se caracteriza principalmente por su gran versatilidad, flexibilidad, alta oferta de parámetros y algoritmos a escoger en cada fase del procesado de datos de alta resolución, procedentes de MS (Katajamaa y Oresic, 2005). Por sus características, es una herramienta muy recomendada para el tratamiento de datos procedentes de CE-MS.

- R: R es un lenguaje y entorno de programación de acceso gratuito empleado principalmente en cálculo, estadística y creación de gráficos (http://www.r-project.org/). XCMS es un paquete (o "extensión") encriptado en lenguaje de programación R enfocado al tratamiento de datos procedentes del acomplamiento de técnicas analíticas de separación con MS, descrito por primera vez en 2006 (Smith y col., 2006). mzMatch es otro paquete gratuito de R con el que se pueden comparar tantos grupos de muestras como se desee. Se caracteriza por su rapidez para el tratamiento de un elevado número de muestras. Además de las extensiones mencionadas (XCMS y mzMatch), R presenta un gran número de utilidades de cálculo, estadísticas y de generación de gráficos. El principal inconveniente de R es el de la programación de las funciones, pues a diferencia de otras herramientas informáticas, no hay botones o ventanas que despliegan menús donde el usuario puede elegir las diferentes opciones.

- IDEOM es una herramienta bioinformática muy novedosa (Creek y col., 2012) con la que se puede combinar el uso de los paquetes XCMS y mzMatch anteriormente mencionados seleccionando los parámetros a escoger en un sencilla interfaz de Excel sin la necesidad de aprender a programar en R. Además permite a su vez filtrar los datos obtenidos para reducir el número de falsos positivos, cotejar con estándares, crear librerías adaptadas a las condiciones experimentales escogidas por el usuario e identificar los metabolitos

usando las bases de datos más importantes en Metabolómica (descritas en el apartado 1.2.5), entre otras acciones.

Independientemente del software utilizado, la lista de señales metabólicas obtenida se debe someter a distintos tipos de normalización para evaluar cuál es el idóneo. Los métodos de normalización más comúnmente aplicados (Katajamaa y Oresic, 2007) tienen en cuenta la señal de estándares internos (Sysi-Aho y col., 2007; Redestig y col., 2009) o el área total de los picos obtenidos del cromatograma/electroferograma (Wang y col., 2003; Scholz y col., 2004). La normalización será más efectiva cuando la variación que se refleje entre muestras procedentes del mismo grupo de clasificación de muestras sea menor. En muchos casos en que la variación entre muestras es reducida, se ha demostrado que es más efectivo no normalizar (Laiakis y col., 2010; Ghannoum y col., 2013) ya que toda normalización implica una modificación de los datos. Por este motivo, en casos en los que sea posible (bajo %RSD) es preferible no normalizar.

1.2.4. Análisis estadístico

Los principales métodos estadísticos aplicados en el área de la Metabolómica han sido publicados recientemente en una interesante revisión bibliográfica (Korman y col., 2012). En este apartado se explicarán técnicas estadísticas ampliamente utilizadas no exclusivas de la Metabolómica, haciendo especial hincapié en aquellas técnicas estadísticas que se han utilizado en los capítulos de esta Tesis.

El análisis estadístico en Metabolómica se realiza generalmente sobre los datos contenidos en una tabla formada por los niveles de múltiples metabolitos (variables) analizados en un conjunto de muestras procedentes de varios grupos. En primer lugar hay que distinguir entre métodos estadísticos "univariantes", que utilizan únicamente la información de cada variable (metabolito), y los "multivariantes" que utilizan de manera conjunta la información de múltiples variables. Puesto que muchos de los métodos estadísticos asumen una distribución normal de los datos y homogeneidad en las varianzas dentro de cada grupo de muestras, se debe realizar un estudio para comprobar dichas premisas, previo a la aplicación de cualquier método para comparar grupos de

muestras. Para comprobar la normalidad de cada variable (los niveles de cada metabolito) en cada grupo, se puede utilizar el test de Kolmogorov-Smirnov (K-S), la corrección de Lilliefors, o el test de Shapiro-Wilk (S-W) que proporcionan niveles de probabilidad de que la variable presente una distribución normal. Niveles de probabilidad mayores a 0.05 permiten aceptar dicha normalidad. Si los datos son normales se aplicarán posteriormente pruebas estadísticas "paramétricas", mientras que si no presentan dicha distribución se realizarán pruebas estadísticas "no paramétricas". Generalmente los test paramétricos comparan valores medios de metabolitos entre grupos mientras que los test no paramétricos contrastan valores de medianas entre grupos. Como la mayoría de los métodos estadísticos paramétricos asumen que los datos poseen valores de varianza homogéneos entre los grupos de muestras (valores de desviación de los metabolitos similar en cada uno de los grupos) es recomendable realizar un estudio previo de dicha homogeneidad. Para comprobar la homogeneidad de las varianzas de las variables en los grupos se puede utilizar el test de Levene que proporciona un nivel de probabilidad. Si dicha probabilidad es mayor de 0.05 permite aceptar la homogeneidad de dichas varianzas.

1.2.4.1. *Análisis estadísticos univariantes*

Una vez que se conoce la distribución y homogeneidad de las variables (niveles de metabolitos) en los grupos de muestras, se puede seleccionar el método de análisis de la varianza más adecuado. El análisis de la varianza es una prueba estadística univariante que tiene como objetivo principal la detección de metabolitos significativamente diferentes entre los grupos con una probabilidad de errar generalmente menor al 0.05, o lo que es lo mismo, un nivel de significación $\alpha < 0.05$. En la Figura 1.4 se muestra un esquema del procedimiento recomendado en la realización de cualquier análisis de varianza. También se muestran las equivalencias aceptadas entre pruebas estadísticas paramétricas y no paramétricas.
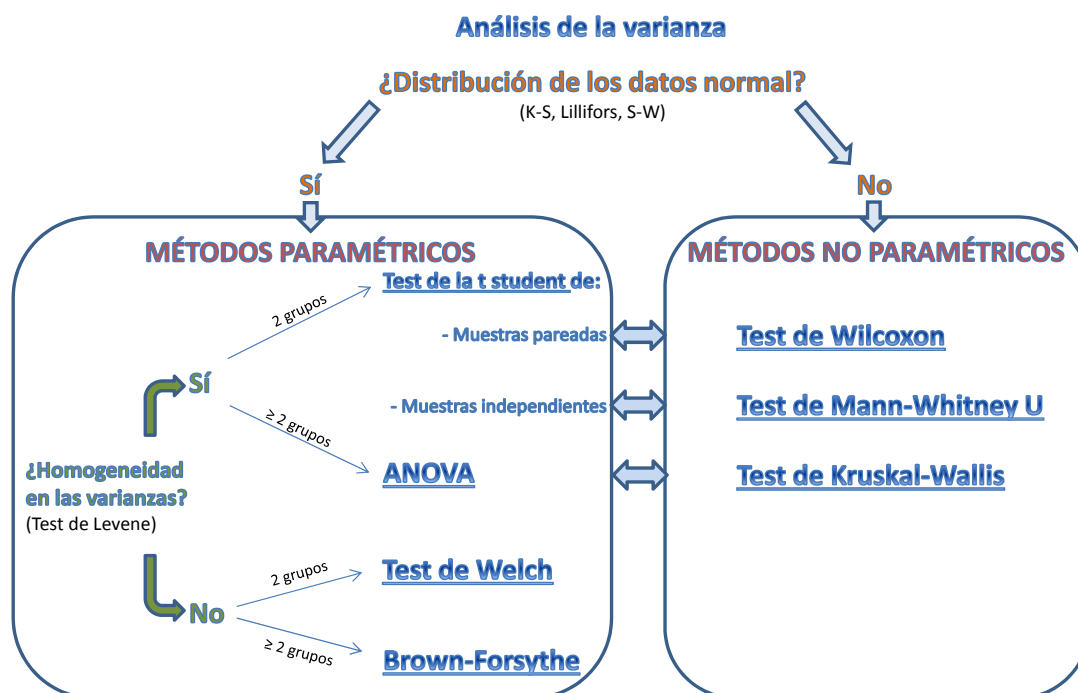
Figura 1.4. Esquema del procedimiento a seguir en un análisis de varianza. Las flechas horizontales con doble punta muestran las equivalencias entre pruebas paramétricas y no paramétricas.

Dentro de los análisis de varianza paramétricos con varianzas homogéneas entre los grupos, destacan por un lado la prueba de la "t de *student*" (*t-test*) clásica cuando se comparan dos grupos de muestras, y por otro lado, el análisis de varianza de una vía ANOVA si se trata de comparar más de dos grupos de muestras. ANOVA es el método estadístico univariante más utilizado, que tiene como objetivo principal contrastar la igualdad de las medias poblacionales de la variable analizada (el metabolito) en las muestras de los distintos grupos. Este método acepta distribuciones normales (prueba paramétrica) y homogeneidad de las varianzas de las variables en dichos grupos. En los capítulos de esta Tesis se ha utilizado ANOVA para la detección de metabolitos significativamente diferentes entre los grupos de muestras, considerando un nivel de significación $\alpha < 0.05$ para el estadístico de contraste F de Snedecor. Si aplicando el ANOVA se observan diferencias, se pueden utilizar diferentes pruebas estadísticas para estudiar en detalle cómo son dichas diferencias entre los grupos de muestras, como el test de la mínima diferencia significativa (*Least Significant Difference test* o LSD), o el test de Bonferroni. Para realizar el análisis de la varianza en el caso de datos normales pero con varianzas no homogéneas entre grupos (datos heterocedásticos) se puede utilizar el test de Welch (para 2 grupos de muestras) o el de Brown Forsythe (≥ 2 grupos de muestras) para contrastar la igualdad de los valores medios (Figura 1.4). Por último, si no es posible aceptar que los datos son normales se pueden aplicar técnicas no

paramétricas, como el test de Mann-Withney o Kruskal-Wallis, para el contraste de la igualdad de las medianas entre los grupos.

1.2.4.2. Análisis estadísticos multivariantes

Entre los métodos estadísticos multivariantes se puede hablar de métodos de clasificación "supervisados", que tienen en cuenta la información de la pertenencia de las muestras a los grupos de clasificación, y los métodos "no supervisados" que no utilizan dicha información. Generalmente en primer lugar se emplean métodos estadísticos no supervisados para aplicar después métodos supervisados.

Los métodos no supervisados al no tener en cuenta la clasificación de las muestras, proporcionan información valiosa de la relación entre las muestras (qué muestras se parecen más entre sí o si existen algunas muy dispares con respecto a las demás), y/o de la relación entre las variables (qué metabolitos o señales metabólicas están más correlacionadas entre sí). Estos métodos ofrecen una visión muy global y útil de los datos, y se pueden emplear en la evaluación y optimización de los distintos parámetros del procesado de datos y/o normalización previamente mencionados en el apartado 1.2.3. Algunos ejemplos de técnicas estadísticas no supervisadas son el análisis de conglomerados (o "*clustering*") y el análisis de componentes principales (PCA). Con la utilización de estos métodos no supervisados es posible la detección de muestras atípicas, muy diferentes al resto, también llamadas "*outliers*". La eliminación de dichos *outliers* corresponde al punto de partida de todo análisis estadístico. La detección de estas muestras atípicas es fundamental y previa a la aplicación de un análisis supervisado. Una vez se dispone del total de muestras válidas (sin *outliers*) y los metabolitos apropiados (con baja desviación dentro de su grupo de clasificación), se puede proceder a la aplicación de técnicas estadísticas de clasificación supervisadas. Estos procedimientos resultan de gran ayuda en Metabolómica para descubrir potenciales biomarcadores relacionados con el diagnóstico de patologías, o para medir el efecto de terapias o medidas de prevención en campos como la medicina o la nutrición. El objetivo fundamental de estas técnicas es obtener funciones de clasificación, a partir de la información proporcionada por las variables analizadas en las muestras pertenecientes a los diferentes grupos que forman el llamado "conjunto de

entrenamiento" ("*training data set*"). Este conjunto de muestras se puede utilizar para confirmar que existe diferencia entre los grupos de clasificación, y también para predecir la clasificación de muestras ciegas ("*test data set*"). El análisis lineal discriminante (LDA) es uno de los métodos de clasificación supervisados más empleados en Metabolómica. El LDA se ha utilizado en esta Tesis para elegir, de entre el total de metabolitos analizados, los que tienen mayor capacidad para diferenciar entre los grupos de muestras (potenciales biomarcadores), pudiéndose establecer con estos metabolitos seleccionados, un modelo predictivo para aplicar a otras muestras cuya clasificación se desconoce. Con la información de los potenciales biomarcadores seleccionados mediante el LDA, los resultados que se obtienen de la aplicación de esta técnica son:

- Las funciones de clasificación: se obtiene una función de clasificación para cada grupo de muestras. Estas funciones son combinaciones lineales de las variables analizadas (niveles de los potenciales biomarcadores).
- La tabla de clasificación: cuando se aplica la función de clasificación a cada una de las muestras, éstas quedarán clasificadas en un grupo u otro teniendo en cuenta únicamente los niveles de los metabolitos seleccionados mediante el LDA (los potenciales biomarcadores). En la tabla de clasificación se muestran los porcentajes del número de muestras correctamente clasificadas.

Si se desea realizar el LDA y el número de variables analizadas es muy grande (gran número de metabolitos detectados) se puede utilizar el procedimiento de selección por pasos sucesivos ("*stepwise procedure*"), que proporciona en cada paso la variable (metabolito) que más contribuye a la diferenciación de los grupos. Un método complementario al LDA es el análisis canónico discriminante (CVA) que permite obtener una representación gráfica bidimensional de los resultados de clasificación del LDA. Las muestras se representan en el plano definido por las dos primeras variables canónicas (CV) discriminantes, que son combinaciones lineales de las variables (metabolitos) seleccionadas en el LDA. Estas variables canónicas están calculadas para lograr la máxima separación entre los grupos de clasificación. A parte de la representación gráfica de las muestras, el CVA puede ayudar a interpretar las diferencias entre los grupos de muestras a partir de la matriz de correlación entre las variables originales y las variables canónicas calculadas ("*factor structure matrix*"). En

esta matriz se puede observar por ejemplo, en qué grupo de clasificación es mayor el nivel de un biomarcador y cuánto mayor a los demás grupos, o cuál es la progresión en los niveles de un metabolito entre los distintos grupos de clasificación. Para conocer el poder predictivo del modelo de clasificación generado mediante el LDA, con los potenciales biomarcadores seleccionados, se pueden emplear dos técnicas. Si se dispone de muestras ciegas (muestras con una clasificación desconocida) que no han sido utilizadas en el cálculo de las funciones de clasificación del LDA, se puede aplicar un método de validación externa. Consiste en aplicar las funciones de clasificación del LDA para estas muestras ciegas obteniendo un porcentaje de asignación correcta de las muestras a su grupo de clasificación. Por el contrario, si únicamente se dispone de las muestras del conjunto de entrenamiento, se puede llevar a cabo un método de validación interna. Entre los métodos de validación interna, la "validación cruzada" es uno de los métodos más frecuentemente utilizados. En esta Tesis se ha utilizado concretamente el método conocido como "método de validación cruzada dejando uno fuera" ("*leave-one-out cross-validation*" o LOO-CV). Consiste en no considerar una de las muestras (dejar fuera la muestra del conjunto de entrenamiento) y aplicar el modelo predictivo del LDA calculado con las muestras restantes, para finalmente ver si dicha muestra se clasifica correctamente en su grupo. Dicho procedimiento se repite para todas las muestras. Mediante la aplicación de estos dos procedimientos de validación (externa e interna) se obtiene información de la potencia predictiva del método de clasificación generado mediante el LDA. En términos generales, un resultado de LOO-CV superior al 80% de clasificación correcta se considera satisfactorio. Si los resultados de validación son adecuados, queda demostrada la potencia predictiva del modelo de clasificación generado y por tanto las funciones de clasificación obtenidas con el LDA pueden utilizarse para predecir la clasificación de nuevas muestras.

1.2.5. Identificación de los metabolitos

Una vez llevado a cabo el análisis estadístico, las señales metabólicas (o metabolitos) estadísticamente significativas, empleadas para la clasificación de los grupos (y por lo tanto, potenciales biomarcadores) se identifican para intentar asociar la variación observada en dichas moléculas con una explicación biológica coherente. Para ello, normalmente se llevan a cabo procesos de identificación tentativa. La

identificación de los metabolitos de interés comienza con una búsqueda en bases de datos de metabolitos, algunas de las más empleadas se destacan en la Tabla 1.2. En dichas bases de datos se coteja la masa obtenida a partir del valor de m/z proporcionado por el espectrómetro de masas y las masas teóricas de los metabolitos contenidos en dichas bases de datos. Lógicamente, cuanto mayor sea la exactitud del valor m/z proporcionado por el espectrómetro de masas para un determinado metabolito, mayor será la probabilidad de obtener una correcta identificación, de ahí la necesidad de utilizar espectrómetros de masas de alta resolución que proporcionan valores de m/z con errores de unas pocas partes por millón e incluso inferiores.

Tabla 1.2. Relación de bases de datos más utilizadas en la identificación tentativa de los metabolitos en muestras biológicas.

Base de datos	Información destacada	Referencia
HMDB	Es una base de datos que contiene información detallada sobre los metabolitos del cuerpo humano. A través de ella se puede obtener información química, clínica y bioquímica del compuesto seleccionado, de forma directa o por medio de enlaces con otras bases de datos (KEGG, PubChem, MetaCyc, ChEBI, PDB, Swiss-Prot, y GenBank). (www.hmdb.ca).	Wishart y col., 2009
METLIN	Es una base de datos enfocada a la espectrometría de masas de metabolitos. Está diseñada por el Centro Scripps para la Espectrometría de Masas. Contiene más de 75000 estructuras. Ofrece información sobre fórmula molecular y estructura de los metabolitos (espectros de masas y espectros de fragmentación), además del enlace a otras bases de datos (KEGG, HMDB, PubChem, LipidMaps). (metlin.scripps.edu/).	Smith y col., 2005
PubChem	Es una base de datos de compuestos químicos. El Centro Nacional de Información Biotecnológica (NCBI) de Estados Unidos se encarga de su actualización y mantenimiento. Contiene la descripción e información estructural de millones de compuestos. La mayoría de moléculas listadas en PubChem tienen un peso molecular inferior a 2000 Da. Dentro de esta base de datos se encuentran tres bases de datos primarias: de compuestos (puros y mezclas), de sustancias (mezclas, extractos y sustancias no caracterizadas) y bioensayos (artículos referentes a bioactividades de moléculas investigadas). Contiene enlaces a más de 20 bases de datos adicionales. (pubchem.ncbi.nlm.nih.gov).	Bolton y col., 2008.
LipidMaps	Es una base de datos enfocada a las sustancias de carácter lipídico. Ofrece información sobre su clasificación, nomenclatura y estructura. Contiene dos bases de datos primarias, una con información estructural (LMSD) y la otra con asociación de los lípidos a proteínas (LMPD). Contiene enlaces con otras bases de datos relacionadas (Cell Migration Gateway, Functional Glycomics Gateway, Omics Gateway, Pathway Interaction Database, RNAi Gateway, Signaling Gateway, Structural Genomics Knowledgebase). (www.lipidmaps.org).	Sud y col., 2007.
KEGG	La Enciclopedia de Genes y Genomas de Kyoto (KEGG) compila los datos sobre sustancias químicas y reactividad de compuestos relacionados con los seres vivos de cinco bases de datos distintas: Compound, Glycan, Reaction, Rpair, y Enzyme. Permite enlazar con más de 20 bases de datos. (www.genome.jp/ligand/)	Kanehisa, 1997
MassBank	Es una base de datos focalizada en el espectro de masas de alta resolución de metabolitos. Se pueden buscar los datos obtenidos de relaciones m/z así como obtener el espectro de la sustancia elegida. Entre las bases de datos a las que se puede acceder a través de ella, se encuentran KEGG, KNapSAcK y LipidBank. (www.massbank.jp)	Horai y col., 2010

Esta comparación de valores de m/z se completa con la comparación de los perfiles isotópicos experimentales y teóricos. Los perfiles isotópicos resultan de gran utilidad debido a que los átomos que forman las moléculas poseen isótopos naturales

minoritarios (en un porcentaje distinto y característico para cada átomo), lo cual queda reflejado en el espectro de masas. La mayoría de los programas de adquisición de MS incluyen la comparación de patrones isotópicos obtenidos experimentalmente con los teóricos, generando una lista de posibles fórmulas moleculares. Integrando dicha información con la obtenida a partir de las bases de datos anteriormente mencionadas se puede realizar una identificación tentativa de los metabolitos. Además las bases de datos están incorporando los patrones de fragmentación MS/MS para su comparación con los obtenidos experimentalmente. A modo de confirmación de la identificación tentativa de los metabolitos obtenida mediante el empleo de las bases de datos, se pueden emplear la inyección y análisis de patrones de aquellos metabolitos que estén disponibles comercialmente y que lógicamente deberán ser analizados en idénticas condiciones a las empleadas con las muestras del estudio metabolómico.

1.2.6.   Interpretación biológica

Una vez obtenida la identificación tentativa de las moléculas de interés (potenciales biomarcadores), es muy interesante vincular su alteración (debida por ejemplo a un proceso fisiopatológico o una intervención dietética) con las rutas metabólicas en las que participan dichos metabolitos. Para ello, estos metabolitos se localizan en las distintas rutas metabólicas. Esta información está contenida en bases de datos como KEGG (Kanehisa, 1997). La mayoría de bases de datos de identificación en metabolómica contienen enlaces a la información de las rutas metabólicas de KEGG por lo que se puede consultar de manera simultánea a la identificación de los metabolitos. Una herramienta muy adecuada para situar todos los metabolitos significativos de un determinado estudio en sus correspondientes rutas metabólicas es MassTRIX (Suhre y col., 2008). Dicha herramienta bioinformática permite, mediante la información de la masa exacta experimental de los metabolitos, situarlos en los mapas de rutas bioquímicas de la base de datos de KEGG. Con la información obtenida, se pueden formular hipótesis del posible efecto/alteración observado entre muestras profundizando en los procesos bioquímicos. Dicha averiguación puede dar lugar a hipótesis o trabajos futuros dirigidos hacia la determinación y cuantificación de metabolitos de especial interés en determinadas rutas metabólicas, o biomarcadores de una determinada enfermedad, proceso, alteración, etc.

REFERENCIAS

Adamko D.J., Sykes B.D., Rowe B.H.; *Chest J* 141 (2012) 1295-1302.

Aharoni A., Ric de Vos C.H., Verhoeven H.A., Maliepaard C.A., Kruppa G., Bino R., Goodenowe D.B.; *Omics* 6 (2002) 217-234.

Altelaar A.F., Munoz J., Heck A.J.; *Nat Rev Genet* 14 (2013) 35-48.

Amorini A.M., Giorlandino C., Longo S., D'Urso S., Mesoraca A., Santoro M.L., Picardi M., Gullotta S., Cignini P., Lazzarino D., Lazzarino G., Tavazzi B.; *Mol Cell Biochem* 359 (2012) 205-216.

Bains, W.; *Trends Biotechnol* 14 (1996) 312-317.

Barding G.A.Jr., Salditos R., Larive C.K.; *Anal Bioanal Chem* 404 (2012) 1165-1179.

Bellew M., Coram M., Fitzgibbon M., Igra M., Randolph T., Wang P., May D., Eng J., Fang R., Lin C., Chen J., Goodlett D., Whiteaker J., Paulovich A., McIntosh M., *Bioinformatics* 22 (2006) 1902-1909.

Birungi G., Chen S.M., Loy B.P., Ng M.L., Li S.F.; *J Proteome Res* 9 (2010) 6523-6534.

Bolton E., Wang Y., Thiessen P.A., Bryant S.H. *PubChem: Integrated Platform of Small Molecules and Biological Activities*, IN *Annual Reports in Computational Chemistry.* Wheeler R.A., Spellmeyer D.C. (Eds.) American Chemical Society, Washington DC, 2008, 217-241.

Brindle J.T., Antti H., Holmes E., Tranter G., Nicholson J.K., Bethell H.W., Clarke S., Schofield P.M., McKilligin E., Mosedale D.E., Grainger D.J.; *Nat Med* 8 (2002) 1439-1444.

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

Bu Q., Huang Y., Yan G., Cen X., Zhao Y.L.; *Comb Chem High Throughput Screen* 15 (2012) 266-275.

Chambers A.G., Mellors J.S., Henley W.H., Ramsey J.M.; *Anal Chem* 83 (2011) 842-849.

Chen H., Pan Z., Talaty N., Raftery D., Cooks R.G.; *Rapid Commun Mass Spectrom* 20 (2006) 1577-1584.

Chen L., Zhou L., Chan E.C., Neo J., Beuerman R.W.; *J Proteome Res* 10 (2011) 4876-4882.

Cifuentes A.; *J Chromatogr A* 1216 (2009) 7109.

Claudino W.M., Goncalves P.H., di Leo A., Philip P.A., Sarkar F.H.; *Crit Rev Oncol Hematol* 84 (2012) 1-7.

Cohen L.H., Gusev A.I.; *Anal Bioanal Chem* 373 (2002) 571-586.

Colebatch G., Desbrosses G., Ott T., Krusell L., Montanari O., Kloska S., Kopka J., Udvardi M.K.; *Plant J* 39 (2004) 487-512.

Collino S., Martin F.P.J., Kochhar S., Rezzi S.; *Chimia* 65 (2011) 396-399.

Creek D.J., Jankevics A., Burgess K.E., Breitling R., Barrett M.P.; *Bioinformatics* 28 (2012) 1048-1049.

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

Dunn W.B., Ellis D.I.; *Trends Analyt Chem* 24 (2005) 285-293.

Eckhart A.D., Beebe K., Milburn M.; *Clin Transl Sci* 2012, 5, 285-288.

Edwards J.L., Kennedy R.T.; *Anal Chem* 77 (2005) 2201-2209.

Ewing A.G., Wallingford R.A., Olefirowicz T.M.; *Anal Chem* 61 (1989) 292-303.

Fan T.W., Lorkiewicz P.K., Sellers K., Moseley H.N., Higashi R.M., Lane A.N.; *Pharmacol Ther* 133 (2012) 366-391.

Fiehn O.; *Comp Funct Genomics* 2 (2001) 155-168.

Fiehn O.; *Plant Mol Biol* 48 (2002) 155-171.

Fiehn O., Kind T.; *Methods Mol Biol* 358 (2007) 3-17.

Fiehn O.; *Trends Analyt Chem* 27 (2008) 261-269.

Friedrich N.; *J Endocrinol* 215 (2012) 29-42.

Ganti S., Weiss R.H.; *Urol Oncol* 29 (2011) 551-557.

Ghannoum M.A., Mukherjee P.K., Jurevic R.J., Retuerto M., Brown R.E., Sikaroodi M., Webster-Cyriaque J., Gillevet P.M.; *OMICS* 17 (2013) 5-15.

Goodacre R., Vaidyanathan S., Dunn W.B., Harrigan G.G., Kell D.B.; *Trends Biotechnol* 22 (2004) 245-252.

Goodwin R.J.A.; *J Proteomics* 75 (2012) 4893-4911.

Greer T., Sturm R., Li L.; *J Proteomics* 74 (2011) 2617-2631.

Gullberg J., Jonsson P., Nordström A., Sjöström M., Moritz T.; *Anal Biochem* 331

(2004) 283-295.

Gygi S.P., Rochon Y., Franza B.R., Aebersold R.; *Mol Cell Biol* 19 (1999) 1720-1730.

Harris G.A., Galhena A.S., Fernández F.M.; *Anal Chem* 83 (2011) 4508-4538.

Hassanein M., Callison J.C., Callaway-Lane C., Aldrich M.C., Grogan E.L., Massion P.P.; *Cancer Prev Res* 5 (2012) 992-1006.

Herrero M., García-Cañas V., Simó C., Cifuentes A.; *Electrophoresis* 31 (2010) 205-228.

Herrero M., Simó C., García-Cañas V., Ibáñez E., Cifuentes A.; *Mass Spectrom Rev* 31 (2011) 49-69.

Hommerson P., Khan A.M., de Jong G.J., Somsen G.W.; *Mass Spectrom Rev* 30 (2011) 1096-1120.

Horai H., Arita M., Kanaya S., Nihei Y., Ikeda T., Suwa K., Ojima Y., Tanaka K., Tanaka S., Aoshima K., Oda Y., Kakazu Y., Kusano M., Tohge T., Matsuda F., Sawada Y., Hirai M.Y., Nakanishi H., Ikeda K., Akimoto N., Maoka T., Takahashi H., Ara T., Sakurai N., Suzuki H., Shibata D., Neumann S., Iida T., Tanaka K., Funatsu K., Matsuura F., Soga T., Taguchi R., Saito K., Nishioka T.; *J Mass Spectrom* 45 (2010) 703-714.

Huang M.Z., Cheng S.C., Cho Y.T., Shie J.; *Anal Chim Acta* 702 (2011) 1-15.

Huck C.W., Bakry R., Huber L.A., Bonn G.K.; *Electrophoresis* 27 (2006) 2063-2074.

Issaq H.J., Van Q.N., Waybright T.J., Muschik G.M., Veenstra T.D.; *J Sep Sci* 32 (2009) 2183-2199.

Jacobs D.M., Fuhrmann J.C., van Dorsten F.A., Rein D., Peters S., van Velzen E.J.J., Hollebrands B., Draijer R., van Duynhoven J., Garczarek U.; *J Agr Food Chem* 60 (2012) 3078-3085.

Jansson J., Willing B., Lucio M., Fekete A., Dicksved J., Halfvarson J., Tysk C., Schmitt-Kopplin P.; *PLoS One* 4 (2009) e6386.

Jiye A., Trygg J., Gullberg J., Johansson A.I., Jonsson P., Antti H., Marklund S.L., Moritz T.; *Anal Chem* 77 (2005) 8086-8094.

Jones D.P., Park Y., Ziegler T.R.; *Annu Rev Nutr* 32 (2012) 183-202.

Kanehisa M.; *Trends Genet* 13 (1997) 375-376.

Katajamaa M., Oresic M.; *BMC Bioinformatics* 6 (2005) 179.

Katajamaa M., Oresic M.; *J Chromatogr A* 1158 (2007) 318-328.

Kell D.B., Brown M., Davey H.M., Dunn W.B., Spasic I., Oliver S.G.; *Nat Rev Microbiol* 3 (2005) 557-565.

Kemperman R.F., Horvatovich P.L., Hoekman B., Reijmers T.H., Muskiet F.A., Bischoff R.; *J Proteome Res* 6 (2007) 194-206.

Kind T., Tolstikov V., Fiehn O., Weiss R.H.; *Anal Biochem* 363 (2007) 185-195.

Kivilompolo M., Pol J., Hyotylainen T.; *LC GC Europe* 5 (2011) 232-243.

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

Kussmann M., Raymond F., Affolter M.; *J Biotechnol* 124 (2006) 758-787.

Laiakis E.C., Morris G.A., Fornace A.J., Howie S.R.; *PLoS One* 5 (2010) e12655.

Lavine B.,Workman J.J.Jr.; *Anal Chem* 76 (2004) 3365-3371.

Law W.S., Huang P.Y., Ong E.S., Ong C.N., Li S.F., Pasikanti K.K., Chan E.C.; *Rapid Commun Mass Spectrom* 22 (2008) 2436-2446.

Lee R., Britz-McKibbin P.; *Anal Chem* 81 (2009) 7047-7056.

Lee Y.J., Perdian D.C., Song Z., Yeung E.S., Nikolau B.J.; *Plant J* 70 (2012) 81-95.

Lei Z., Huhman D.V., Sumner L.W.; *J Biol Chem* 286 (2011) 25435-25442.

Llorach R., Garcia-Aloy M., Tulipani S., Vazquez-Fresno R., Andres-Lacueva C.; *J Agric Food Chem* 60 (2012) 8797-8808.

Martin F.P.J., Rezzi S., Pere-Trepat E., Kamlage B., Collino S., Leibold E., Kastler J., Rein D., Fay L.B., Kochhar S.; *J Proteome Res* 8 (2009) 5568-5579.

McGhie T.K., Rowan D.D.; *Mol Nutr Food Res* 56 (2012) 147-158.

McNiven E.M., German J.B., Slupsky C.M.; *J Nutr Biochem* 22 (2011) 995-1002.

Mehrotra B., Mendes P. *Bioinformatics approaches to integrate metabolomics and other systems biology data*, IN *Plant Metabolomics*. Saito K., Dixon R.A., Willmitzer L. (Eds.) Springer-Verlag. Berlin, 2006, 105-115.

Mellert W., Kapp M., Strauss V., Wiemer J., Kamp H., Walk T., Looser R., Prokoudine A., Fabian E., Krennrich G., Herold M., van Ravenzwaay B.; *Toxicol Lett* 207 (2011)

173-181.

Michalke B.; *Electrophoresis* 26 (2005) 1584-1597.

Michopoulos F., Lai L., Gika H., Theodoridis G., Wilson I.; *J Proteome Res* 8 (2009) 2114-2121.

Mitrevski B.S., Kouremenos K.A., Marriott P.J.; *Bioanalysis* 1 (2009) 367-391.

Miura D., Fujimura Y., Tachibana H., Wariishi H.; *Anal Chem* 82 (2010) 498-504.

Miura D., Fujimura Y., Wariishi H.; *J Proteomics* 75 (2012) 5052-5060.

Moini M.; *Anal Chem* 79 (2007) 4241-4246.

Mondello L., Tranchida P.Q., Dugo P., Dugo G.; *Mass Spectrom Rev* 27 (2008) 101-124.

Mostafa A., Edwards M., Górecki T.; *J Chromatogr A* 1255 (2012) 38-55.

Murray J.A.; *J Chromatogr A* 1261 (2012) 58-68.

Nambiar P.R., Gupta R.R., Misra V.; *Mutat Res* 693 (2010) 3-18.

Nicholson J.K., Lindon J. C., Holmes E.; *Xenobiotica* 29 (1999) 1181-1189.

Novotná H., Kmiecik O., Gałązka M., Krtková V., Hurajová A., Schulzová V., Hallmann E., Rembiałkowska E., Hajšlová J.; *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* 29 (2012) 1335-1346.

Olivares J.A., Nguyen N.T., Yonker C.R., Smith R.D.; *Anal Chem* 59 (1987) 1230-1232.

Oliver S.G., Winson M.K., Kell D.B., Baganz F.; *Trends Biotechnol* 16 (1998) 373-378.

Pasikanti K.K., Ho P.C., Chan E.C.; *J Chromatogr B Analyt Technol Biomed Life Sci* 871 (2008) 202-211.

Patti G.J.; *J Sep Sci* 34 (2011) 3460-3469.

Pereira H., Martin J.F., Joly C., Sebedio J.L, Pujos-Guillot E.; *Metabolomics* 6 (2010) 207-218.

Pierce K.M., Hoggard J.C., Mohler R.E., Synovec R.E.; *J Chromatogr A* 1184 (2008) 341-352.

Pol J., Hyötyläinen T.; *Anal Bioanal Chem* 391 (2008) 21-31.

Polson C., Sarkar P., Incledon B., Raguvaran V., Grant R.; *J Chromatogr B Anal Tech Biomed Life Sci* 785 (2003) 263-275.

Puiggròs F., Solà R., Bladé C., Salvadó M.J., Arola L.; *J Chromatogr A* 1218 (2011) 7399-7414.

Putri S.P., Yamamoto S., Tsugawa H., Fukusaki E.; *J Biosci Bioeng* (2013) (doi: 10.1016/j.jbiosc.2013.01.004). In press.

Ramautar R., Somsen G.W., de Jong G.J.; *Electrophoresis* 30 (2009) 276-291.

Ramautar R., Mayboroda O.A., Somsen G.W., de Jong G.J.; *Electrophoresis* 32 (2011) 52-65.

Ramautar R., Somsen G.W., de Jong G.J.; *Electrophoresis* 34 (2013) 86-98.

Redestig H., Fukushima A., Stenlund H., Moritz T., Arita M., Saito K., Kusano M.; *Anal Chem* 81 (2009) 7974-7980.

Rezzi S., Vera F.A., Martin F.P., Wang S., Lawler D., Kochhar S.; *J Chromatogr B Anal Tech Biomed Life Sci* 871 (2008) 271-278.

Robertson, D.G.; *Toxicol Sci* 85 (2005) 809-822.

Roessner U., Wagner C., Kopka J., Trethewey R.N., Willmitzer L.; *Plant J* 23 (2000) 131-142.

Rosenling T., Stoop M.P., Smolinska A., Muilwijk B., Coulier L., Shi S., Dane A., Christin C., Suits F., Horvatovich P.L., Wijmenga S.S., Buydens L.M., Vreeken R., Hankemeier T., van Gool A.J., Luider T.M., Bischoff R.; *Clin Chem* 57 (2011) 1703-1711.

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

Ryan D., Robards K., Prenzler P.D., Kendall M.; *Anal Chim Acta* 684 (2011) 17-29.

Saric J., Want E.J., Duthaler U., Lewis M., Keiser J., Shockcor J.P., Ross G.A., Nicholson J.K., Holmes E., Tavares M.F.; *Anal Chem* 84 (2012) 6963-6972.

Senn T., Hazen S.L., Tang W.H.; *Prog Cardiovasc Dis* 55 (2012) 70-76.

Scholz M., Gatzek S., Sterling A., Fiehn O., Selbig J.; *Bioinformatics* 20 (2004) 2447- 2454.

Shah S.H., Kraus W.E., Newgard C.B.; *Circulation* 126 (2012) 1110-1120.

Smith R.D., Barinaga C.J., Udseth H.R.; *Anal Chem* 60 (1988)1948-1952.

Smith C.A., O'Maille G., Want E.J., Qin C., Trauger S.A., Brandon T.R., Custodio D.E., Abagyan R., Siuzdak G.; *Ther Drug Monit* 27 (2005) 747-751.

Smith C.A., Want E.J., O'Maille G., Abagyan R., Siuzdak G.; *Anal Chem* 78 (2006) 779-787.

Smolinska A., Blanchet L., Buydens L.M., Wijmenga S.S.; *Anal Chim Acta* 750 (2012) 82-97.

Snowden S., Dahlén S.E., Wheelock C.E.; *Bioanalysis* 4 (2012) 2265-2290.

Song E.J., Babar S.M., Oh E., Hasan M.N., Hong H.M., Yoo Y.S.; *Electrophoresis* 29 (2008) 129-142.

Sud M., Fahy E., Cotter D., Brown A., Dennis E.A., Glass C.K., Merrill A.H.Jr., Murphy R.C., Raetz C.R., Russell D.W., Subramaniam S.; *Nucleic Acids Res* 35 (2007) D527-D532.

Sugimoto M., Hirayama A., Ishikawa T., Robert M., Baran R., Uehara K., Kawai K., Soga T., Tomita M.; *Metabolomics* 6 (2010) 27-41.

Sugimoto M., Kawakami M., Robert M., Soga T., Tomita M.; *Curr Bioinform* 7 (2012) 96-108.

Suhre K., Schmitt-Kopplin P.; *Nucleic Acids Res* 36 (2008) W481-W484.

Suhre K., Meisinger C., Döring A., Altmaier E., Belcredi P., Gieger C., Chang D., Milburn M.V., Gall W.E., Weinberger K.M., Mewes H.W., Hrabé de Angelis M., Wichmann H.E., Kronenberg F., Adamski J., Illig T.; *PLoS One* 5 (2010) e13953.

Sumner L.W., Mendes P., Dixon R.A.; *Phytochemistry* 62 (2003) 817-836.

Svatos A.; *Trends Biotechnol* 28 (2010) 425-434.

Syggelou A., Iacovidou N., Atzori L., Xanthos T., Fanos V.; *Pediatr Clin North Am* 59 (2012) 1039-1058.

Sysi-Aho M., Katajamaa M., Yetukuri L., Oresic M.; *BMC Bioinformatics* 8 (2007) 93.

Takats Z., Wiseman J.M., Gologan B., Cooks R.G.; *Science* 306 (2004) 471-473.

Tautenhahn R., Bottcher C., Neumann S.; *BMC Bioinformatics* 9 (2008) 504.

Temmerman L., De Livera A.M., Bowne J.B., Sheedy J.R., Callahan D.L., Nahid A., De Souza D.P., Schoofs L., Tull D.L., McConville M.J., Roessner U., Wentworth J.M.; *Diabetes Metab* 6 (2012) S6.

Theodoridis G.A., Gika H.G., Want E.J., Wilson I.D.; *Anal Chim Acta* 711 (2012) 7-16.

Tiziani S., Emwas A.H., Lodi A., Ludwig C., Bunce C.M., Viant M.R., Günther U.L.; *Anal Biochem* 377 (2008) 16-23.

Vaidyanathan S., Jones D., Broadhurst D.I., Ellis J., Jenkins T., Dunn W.B., Hayes A., Burton N., Oliver S.G., Kell D.B.; *Metabolomics* 1 (2005) 243-250.

Vaidyanathan S., Gaskell S., Goodacre R.; *Rapid Commun Mass Spectrom* 20 (2006) 1192-1198.

Vaidyanathan S., Jones D., Ellis J., Jenkins T., Chong C., Anderson M., Goodacre R.; *Rapid Commun Mass Spectrom* 21 (2007) 2157-2166.

Van Dorsten F.A., Grun C.H., Van Velzen E.J.J., Jacobs D.M., Draijer R., Van Duynhoven J.P.M.; *Mol Nutr Food Res* 54 (2010) 897-908.

Van Kampen J.J., Burgers P.C., de Groot R., Gruters R.A., Luider T.M.; *Mass Spectrom Rev* 30 (2011) 101-120.

Venter A., Nefliu M., Cooks R.G.; *Trends Analyt Chem* 27 (2008) 284-290.

Vuckovic D.; *Anal Bioanal Chem* 403 (2012) 1523-1548.

Wang W., Zhou H., Lin H., Roy S., Shaler T. A., Hill L. R., Norton S., Kumar P., Anderle M., Becker C. H.; *Anal Chem* 2003, 75, 4818-4826.

Wang Y., Griffiths W.J.; *Curr Anal Chem* 3 (2007) 103-126.

Wang X., Sun H., Zhang A., Wang P., Han Y.; *J Sep Sci* 34 (2011) 3451-3459.

Want E.J., Wilson I.D., Gika H., Theodoridis G., Plumb R.S., Shockcor J., Holmes E., Nicholson J.K.; *Nat Protoc* 5 (2010) 1005-1018.

Weston D.J.; *Analyst* 135 (2010) 661-668.

Whitfield P.D., German A.J., Noble P.J. Br.; *J Nutr* 92 (2004) 549-555.

Wishart D.S., Lewis M.J., Morrissey J.A., Flegel M.D., Jeroncic K., Xiong Y., Cheng D., Eisner R., Gautam B., Tzur D., Sawhney S., Bamforth F., Greiner R., Li L.; *J Chromatogr B Analyt Technol Biomed Life Sci* 871 (2008) 164-173.

Wishart D.S., Knox C., Guo A.C., Eisner R., Young N., Gautam B., Hau D.D., Psychogios N., Dong E., Bouatra S., Mandal R., Sinelnikov I., Xia J., Jia L., Cruz J.A., Lim E., Sobsey C.A., Shrivastava S., Huang P., Liu P., Fang L., Peng J., Fradette R., Cheng D., Tzur D., Clements M., Lewis A., De Souza A., Zuniga A., Dawe M., Xiong

Y., Clive D., Greiner R., Nazyrova A., Shaykhutdinov R., Li L., Vogel H.J., Forsythe I.; *Nucleic Acids Res* 37 (2009) D603-610.

Wong M.C.Y., Lee W.T.K., Wong J.S.Y., Frost G., Lodge J.; *J Chromatogr B* 871 (2008) 341-348.

Yoshida M., Hatano N., Nishiumi S., Irino Y., Izumi Y., Takenawa T., Azuma T.; *J Gastroenterol* 47 (2012) 9-20.

Zhang X., Yap Y., Wei D., Chen G., Chen, F.; *Biotechnol Adv* 26 (2008) 169-176.

Zhang B., Powers R.; *Future Med Chem* 4 (2012) 1273.

Zhang A., Sun H., Wang X.; *Appl Biochem Biotechnol* 168 (2012a) 1718-1727.

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

Zhao S.S., Zhong X., Tie C., Chen D.D.; *Proteomics* 12 (2012) 2991-3012.

2. OBJETIVOS

Esta Tesis doctoral muestra una serie de aplicaciones específicas de la Metabolómica dentro del área de alimentos y de la biomedicina, y se ha dividido en dos secciones. La primera sección contiene tres capítulos dedicados a la aplicación y desarrollo de procedimientos analíticos para el estudio metabolómico del efecto antiproliferativo de un extracto rico en polifenoles procedente de romero (*Rosmarinus officinalis*) en HT-29, una línea tumoral humana derivada de adenocarcinoma de colon. Se seguirá además una aproximación Foodómica, incluyendo la integración de los datos obtenidos principalmente en Metabolómica, junto con resultados de Proteómica y Transcriptómica. La segunda sección comprende cuatro capítulos sobre la aplicación de la Metabolómica para el estudio de la enfermedad de Alzheimer. Se ha llevado a cabo el desarrollo metodológico dirigido a la búsqueda de biomarcadores metabólicos que permitan un diagnóstico precoz de la enfermedad de Alzheimer.

Los objetivos generales de esta Tesis son:

- Desarrollar y optimizar nuevos métodos analíticos avanzados útiles para el análisis metabolómico de fluidos biológicos.

- Integrar los datos obtenidos de múltiples plataformas analíticas para incrementar el espectro de metabolitos detectados.

- Desarrollar las herramientas bioinformáticas más adecuadas para conseguir un correcto procesado de los datos en función de la técnica analítica utilizada y la muestra biológica estudiada, optimizando y validando cada una de las etapas involucradas.

- Diseñar y aplicar el análisis estadístico más apropiado en función de los datos metabolómicos obtenidos.

- Obtener una visión detallada de los mecanismos moleculares alterados, relacionándolos con la actividad antiproliferativa de un extracto rico en polifenoles frente a cultivos celulares modelo de cáncer o con la progresión de la enfermedad de Alzheimer.

- Integrar los resultados metabolómicos con aquellos procedentes de otras técnicas ómicas (Proteómica y Transcriptómica) para obtener una evaluación Foodómica global de las muestras estudiadas.
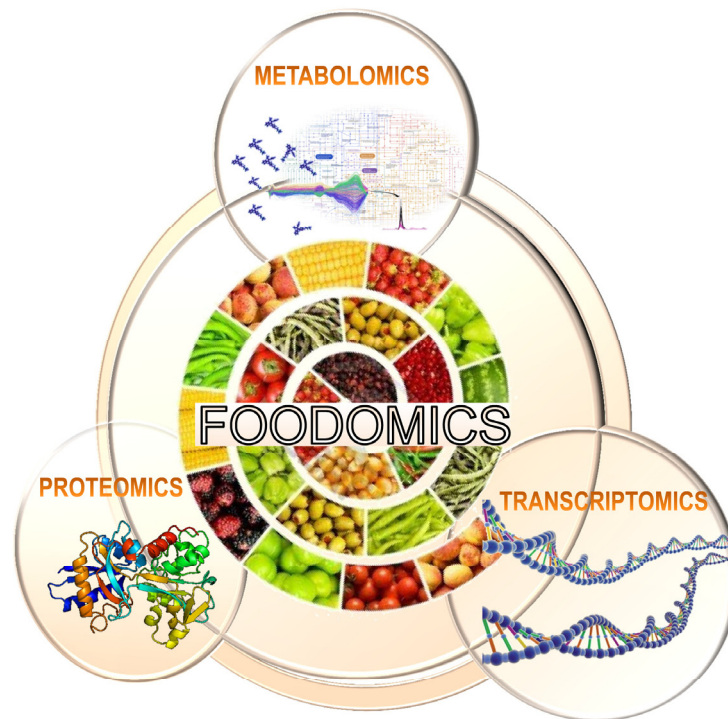
3. PRIMERA PARTE

ESTUDIO METABOLÓMICO DEL EFECTO ANTIPROLIFERATIVO DE INGREDIENTES ALIMENTARIOS EN MODELOS CELULARES HUMANOS DE CÁNCER DE COLON

3.1. INTRODUCCIÓN

La Metabolómica se ha establecido como una herramienta muy útil para profundizar en el conocimiento, la prevención, tratamiento y monitorización terapéutica de múltiples trastornos fisiopatológicos entre los que se encuentra el cáncer (Serkova y Glunde, 2009; Metallo, 2012; Verma y col., 2013). Así, la investigación del cáncer a través del análisis metabolómico ha resultado de enorme utilidad en la detección de fases pretumorales (Montrose y col., 2012), en el descubrimiento de nuevas dianas terapéuticas y en la investigación de los procesos de tumorigénesis (Griffin y Shockcor, 2004). El empleo de la Metabolómica en la investigación del cáncer se remonta a más de 25 años atrás cuando se demostró su gran potencial en la idenficación de tumores malignos a través de un simple análisis de muestras de sangre mediante NMR (Fossel y col., 1986). Actualmente, gracias al enorme desarrollo instrumental, la Metabolómica se ha consolidado como una valiosa herramienta en el avance de la investigación del cáncer tal y como se puede apreciar en el gran número de revisiones bibliográficas publicadas en los últimos años (Kim y col., 2008; Feng y col., 2009; Markuszewski y col., 2010; Nambiar y col., 2010; Wang y col., 2010; Davis y col., 2011; Nagrath y col., 2011; Ng y col., 2011; Bu y col., 2012; O'Connell, 2012; Metallo, 2012; Williams et al 2013). De entre todos los tipos de cáncer, el cáncer de colon o colorrectal (CRC) representa cerca del 10% de todos los tipos de cáncer en el mundo (Stewart y Kleihues, 2003). En la actualidad se diagnostican en España más de 25000 casos nuevos de cáncer de colon al año (Andreu-García y col., 2009), llegando a ser el cáncer con mayor incidencia en nuestro país y la segunda causa de muerte por cáncer en la Unión Europea (Brenner y col., 2012). En la Tabla 3.1 se presentan los estudios metabolómicos relacionados con la investigación en cáncer de colon publicados hasta la fecha (Mayo de 2013), según la base de datos ISI Web of Knowledge.

Tabla 3.1. Artículos en los que se estudia el cáncer de colon mediante estrategias metabolómicas, publicados hasta la fecha (Mayo de 2013) según la base de datos ISI Web of Knowledge

Muestra	Tipo de análisis	Técnica analítica	Procesamiento de datos	Referencia
Cultivos celulares SW1116 y SW480 frente a cultivo de célula sana NCM460	Huella metabólica	GC-TOF MS	AMDIS 2.1	Zimmermann y col., 2007
Tejido Pacientes con CRC primario (n=27) y mucosas control (n=18)	Huella metabólica	GC-TOF MS	ChromaTOF 2.32; R	Denkert y col., 2008
Extractos acuosos fecales Pacientes con CRC (n=21) y sanos (n=11)	Huella metabólica	NMR	TopSpin 1.5; MATLAB; LIBRA	Monleón y col., 2009
Tejido Pacientes con CRC (n=31). Tejido de tumor y control de mucosa	Huella metabólica	HR-MAS NMR GC-TOF MS	MATLAB; SIMCA-P; Shimazu GCMSsolution 2.5	Chan y col., 2009
Tejido Pacientes con CRC (n=16) y pacientes con cáncer de estómago (n=12)	Huella metabólica	CE-TOF MS HPLC-TOF MS	MH; R (XCMS); Multiexperiment Viewer	Hirayama y col., 2009
Suero Pacientes con CRC (n=38), pacientes con adenoma (n=8) y sanos (n=19)	Análisis diana	NMR	NMRLab; PLS Toolbox	Ludwig y col., 2009
Suero Pacientes con CRC (n=31) y sanos (n=8)	Huella metabólica	GC-TOF MS	AMDIS; MATLAB 7.0	Ma y col., 2009a
Orina Pacientes con CRC (n=24) antes y después de la cirugía	Huella metabólica	UPLC-TOF MS	Micromass; MarkerLynx Applications Manager version 4.0; SIMCA-P; SPSS	Ma y col., 2009b
Cultivo celular HT-29 (Tratamiento con inhibidores de HDAC[a] y no HDAC[a])	Análisis diana	Espectrofotometría GC-TOF MS	(sin especificar)	Alcarraz-Vizán y col., 2010
Tejido Pacientes con CRC (n=12): muestra de tumor y de mucosa normal	Huella metabólica	NMR	TopSpin 2.1; Sparky; R; AMIX	Chae y col., 2010
Suero Antes y después de cirugía de 30 pacientes con CRC	Huella metabólica	GC-TOF MS	AMDIS; MATLAB 7.0	Ma y col., 2010
Orina Pacientes con CRC (n=60) y sanos (n=63) **Orina de rata** Antes y después de cirugía de ratas enfermas tratadas con DMH[b] (n=8) y controles sanos(n=8)	Huella metabólica Análisis metabonómico de DMH[b]	GC-TOF MS	XCMS online; MATLAB 7.0; SIMCA-P	Qiu y col., 2010
Suero Procedente de 3 centros. Pacientes con CRC (n=112), controles (n=110). Grupo de validación: pacientes con CRC (n=110) y controles (n=110)	Huella metabólica	FT-ICR-MS HPLC-Q/TOF MS	XMASS; DISCOVA-metrics; ABI QSTAR XL; JMP 8.0.1; SAS 9.2; R 2.9.0; JROCFIT	Ritchie y col., 2010
Tejido Tumoral y no tumoral de 31 pacientes con MSI[c] (n=18) y MSS[d] (n=14) y de individuos sanos (n=31)	Huella metabólica	¹H HR MAS NMR	Matlab	Tessem y col., 2010

Muestra / Descripción	Aplicación	Técnica	Software	Referencia
Suero Pacientes con CRC (n=42) y sanos (n=8)	Perfil de ácidos grasos	GC-TOF MS	(sin especificar)	Kondo y col., 2011
Cultivo celular SW480 y SW620	Perfil del metabolismo energético	NMR	Chenomx NMR Suite 4.6; The Unscrambler9.7	Maddula y Baumbach, 2011
Cultivo celular HT-29 (cuatro purificaciones de metabolitos distintas)	Huella metabólica	CE-TOF MS	DataAnalysis 4.0	Simó y col., 2011
Orina Sanos (n=444) y enfermos de CRC (n=116)	Huella metabólica	NMR	Chenomx NMRsuite v7.0; SIMCAP; STATA/SE 10.1	Wang y col., 2011
Suero Pacientes con CRC metastásico (n=153) y sanos (n=139)	Huella metabólica	NMR	(sin especificar)	Bertini y col., 2012
Orina Pacientes con CRC (n=61) y sanos (n=62) Grupo de validación: 40 pacientes con CRC y 41 sanos	Huella metabólica	GC-TOF MS UPLC-Q/TOF MS	SIMCA-P	Cheng y col., 2012
Cultivo celular HT-29 y SW480 (tratamiento con extracto de aceite de oliva)	Análisis diana de metabolitos fenólicos	nanoLC-TOF MS	DataAnalysis 4.0	Fernández-Arroyo y col., 2012
Cultivo celular HT-29 (tratamiento con extracto polifenólico)	Huella metabólica	CE-TOF MS UPLC-TOF MS	R (XCMS); MH, MPP	Ibañez y col., 2012a
Cultivo celular HT-29 (tratamiento con extracto polifenólico)	Huella metabólica	CE-TOF MS UPLC-TOF MS	DataAnalysis 4.0; R (XCMS); MH, MPP	Ibañez y col., 2012b
Suero Pacientes con CRC (n=59) y controles (n=58)	Perfil de aminoácidos	FIA[e]-MS	Analyst 1.4.2; R	Leichtle y col., 2012
Tejido Pacientes con CRC (n=31): muestras de tumor y de mucosa normal	Huella metabólica	GC x GC-TOF MS	ChromaTOF 4.21; SIMCA-P; GraphPad Prism; SPSS	Mal y col., 2012
Tejido, Plasma y Heces de ratones Muestras recogidas en momentos distintos: 3, 5 y 7 semanas tras administración de azoxymethane (n=40) o suero salino (n=35).	Huella metabólica	UPLC/MS/MS GC-TOF MS	Metabolon's software	Montrose y col., 2012
Suero Pacientes con CRC (n= 60; 12 en cada una de las 4 fases de cáncer) y sanos (n=60). (Grupo de validación 59 pacientes con CRC y 63 voluntarios sanos)	Perfil metabólico	GC-TOF MS	MetAlign; Aloutput; JMP9	Nishiumi y col., 2012
Tejido de ratones (n=131) con y sin mutación en gen APC **Suero de ratones** (n=126) con y sin mutación en gen APC **Cultivos celulares SW480** (n=75) expresando parcial/totalmente el gen APC	Huella metabólica	GC-TOF MS	MetAlign; Aloutput	Yoshie y col., 2012
Tejido Mucosa de tejido tumoral (n=83) y sano (n=87) procedentes de 26 pacientes con CRC.	Huella metabólica	NMR	TopSpin 2.2; Amix 3.9.9; MATLAB;SIMCA-P	Jiménez y col., 2013
Suero Pacientes con CRC (n=52) y controles (n=52).	Perfil de lípidos	DI-ESI(+/-)FT-ICR MS	ApexControl 3.0.0 DataAnalysis; DAssis; SIMCA-P; SPSS	Li y col., 2013
Tejido Pacientes con CRC recurrente (n=4) y no recurrente (n=3) tras tratamiento con 5-fluorouracilo	Perfil de nucleótidos y derivados	GC x GC-TOF MS	SIMCA-P	Phua y col., 2013

ª HDAC: Histone deacetylase; ᵇ DMH: 1,2-dimethylhydrazine; ᶜ MSI: Inestabilidad microsatellite;ᵈ MSS: Estabilidad microsatellite; ᵉ FIA: Análisis por inyección en flujo.

En la Tabla 3.1. se puede observar que el empleo de estrategias metabolómicas en la investigación de CRC es de reciente aparición, en concreto el primer trabajo se publicó en el año 2007 (Zimmermann y col., 2007). Asimismo se aprecia el análisis predominante de muestras de tejido intestinal y un enfoque mayoritario hacia la obtención de la huella metabólica. En el caso del análisis de cultivos celulares, la aproximación metabólica más empleada es también la de la obtención de la ya mencionada huella metabólica, en estudios de comparación de especímenes (fluidos biológicos, tejidos) con y sin CRC, o bien, en la determinación de los efectos a nivel metabolómico de fármacos, xenobióticos e ingredientes alimentarios. Aunque los modelos *in vitro* son una simplificación de una realidad mucho más compleja y la información que se obtiene es en ocasiones parcial, los modelos de cultivos de células tumorales son herramientas muy valiosas a la hora de analizar ciertos aspectos de la biología celular del cáncer de cólon, reduciendo además significativamente la variabilidad entre replicados biológicos. Desde hace algunos años, la relación entre cáncer y dieta es objeto de multitud de estudios (Mosby y col., 2012; Tetè y col., 2012; Thomson, 2012). Existe un número creciente de estudios relativos al efecto beneficioso de ciertos componentes de la dieta contra el desarrollo de cáncer colorrectal, como los polifenoles (Rudolf y col., 2007; Sant y col., 2007; Araujo y col., 2011; Bobe y col., 2012), las isoflavonas (Bennink, 2001), los fitoesteroles (Rao y Janezic, 1992), los ácidos grasos (Williams y col., 2010; Cai y col., 2012; Gerber, 2012; Key y col., 2012; Wu y col., 2012), la fibra (Hansen y col., 2012), las proteínas (Williams y col., 2010), los carbohidratos (Aune y col., 2012), los carotenoides (Wang y col., 2012), las vitaminas (Gorham y col., 2005; Key y col., 2012; Tavani y col., 2012) y los minerales (Chen y col., 2012; Key y col., 2012; Wang y col., 2012; Wark y col., 2012). Una dieta adecuada junto con otras medidas de promoción de la salud (p. ej. un estilo de vida saludable) adquieren una gran relevancia en enfermos con CRC, ya que las posibilidades de curación cuando la enfermedad se diagnostica a tiempo y el tumor es pequeño son muy altas, debido al lento desarrollo de la misma. Se han realizado estudios epidemiológicos en los que se ha asociado una alta ingesta de carne roja y/o alcohol con la aparición de este tipo de cáncer, considerándose por el contrario un factor de protección, el consumo de fibra (Huxley y col., 2009). En este sentido, se ha observado una menor incidencia en poblaciones que consumen preferentemente una dieta de "estilo mediterráneo" (Hadziabdić y col., 2012). Con el fin de entender el efecto que la dieta tiene sobre la salud, es necesario un profundo estudio de los

mecanismos de acción a nivel molecular de los nutrientes y de los constituyentes bioactivos de los alimentos. El número de trabajos que demuestran que ciertos componentes de los alimentos pueden regular la expresión génica de diferentes maneras (Mead, 2007) se encuentra en continuo crecimiento. La Nutrigenómica, una importante disciplina dentro del marco global de la Foodómica, se centra en el estudio del efecto sobre la salud de determinados compuestos de los alimentos o dietas, a través del estudio integrado de la información obtenida mediante las diferentes tecnologías ómicas. Además del análisis metabolómico, el análisis a nivel transcriptómico y proteómico, así como la integración de la información obtenida mediante estas tecnologías ómicas desde una perspectiva de Biología de Sistemas, es probablemente uno de los grandes retos en la investigación hasta alcanzar lo que se denomina "dieta personalizada". Existen todavía muchos interrogantes sobre el efecto de los componentes de la dieta sobre la prevención y el riesgo de padecer cáncer. En este sentido, las líneas celulares tumorales han demostrado su utilidad en la investigación de los efectos de ciertos componentes de la dieta al realizarse en unas condiciones altamente controladas, hecho de especial importancia en la investigación de CRC donde se estima que más de la mitad de los casos de cáncer de colon diagnosticados están relacionados con la dieta y el estilo de vida (Huxley y col., 2013).

La **Primera Parte** de esta Tesis comprende tres capítulos en los que se describe la optimización y el desarrollo de los métodos de análisis para el estudio metabolómico de cultivos celulares modelo de cáncer de colon. La metodología desarrollada se ha aplicado al estudio del efecto de un extracto rico en polifenoles obtenido a partir de romero sobre la línea celular HT-29, derivada de células de adenocarcinoma de colon humano. Este extracto se obtuvo en el laboratorio de Foodómica mediante métodos de extracción con fluidos presurizados medioambientalmente limpios (Herrero y col., 2010). Concretamente, la extracción se llevó a cabo empleando CO_2 supercrítico con un 7% de etanol como modificador polar, a una temperatura de 40ºC y 150 bares de presión. Este extracto de romero demostró tener propiedades antioxidantes (EC_{50}=8.1 µg/mL) y alto contenido en fenoles (0.121 mg ácido gálico/mg de extracto). La caracterización del extracto mediante UPLC-MS/MS (Herrero y col., 2010) reveló altas concentraciones de ácido carnósico (151.55 µg/mg extracto) y carnosol (226.39 µg/mg extracto). La determinación del perfil transcriptómico y de enriquecimiento funcional de células de cáncer de colon tras el tratamiento con este extracto (Valdés y col., 2013)

mostró que éste presentaba una acusada actividad quimiopreventiva reduciendo significativamente la viabilidad celular. Con el fin de profundizar en los procesos bioquímicos alterados por estos compuestos bioactivos se llevó cabo un estudio metabolómico e identificar así las rutas metabólicas afectadas con el objetivo de profundizar en el mecanismo de acción y determinar posibles biomarcadores del efecto antiproliferativo de los polifenoles procedentes del extracto de romero.

En el **Capítulo 1** se ha realizado un estudio comparativo del efecto del tipo de purificación de los metabolitos de las células HT-29 empleadas como modelo humano de cáncer de colon. Mediante CE-TOF MS se ha determinado y comparado la composición y las cantidades relativas de los metabolitos obtenidos tras la aplicación de diversos tipos de purificación de metabolitos antes del estudio metabolómico: extracción en fase sólida (SPE) con dos tipos de material adsorbente, extracción con disolventes orgánicos y ultrafiltración con membranas de corte de 3 kDa.

Los siguientes capítulos (**Capítulo 2** y **Capítulo 3**) están dirigidos a la determinación de los efectos moleculares de un extracto de romero rico en polifenoles con propiedades antiproliferativas sobre células de cáncer de colon HT-29. En el **Capítulo 2** se determinaron las diferencias metabolómicas de células HT-29 cultivadas en presencia de dicho extracto rico en polifenoles, y células control. Con el objetivo de obtener una extensa información metabolómica e identificar posibles biomarcadores de los efectos del extracto, se analizaron e integraron los resultados procedentes de tres plataformas analíticas complementarias, concretamente, CE-TOF MS y UPLC-TOF MS, empleando en este último caso dos modalidades de separación cromatográfica: cromatografía de interacción hidrofílica (HILIC) y fase reversa (RP). Finalmente, en el **Capítulo 3** se llevó a cabo el estudio del efecto del extracto de romero, empleando diferentes tecnologías ómicas (Transcriptómica, Proteómica y Metabolómica) con el fin de obtener una visión general de los procesos celulares modificados, siguiendo una aproximación Foodómica global.

REFERENCIAS

Alcarraz-Vizán G., Boren J., Lee W.N.P., Cascante M.; *Metabolomics* 6 (2010) 229-237.

Andreu-García M., Marzo M., Mascort J., Quintero E., García-Alfonso P., López-Ibor C., Castells A.; *Aten Primaria* 41 (2009) 127-128.

Araujo J.R., Goncalves P., Martel F.; *Nutr Res* 31 (2011) 77-87.

Aune D., Chan D.S.M., Lau R.,Vieira R., Greenwood D.C., Kampman E., Norat T.; *Cancer Cause Control* 23 (2012) 521-535.

Bennink, M.R.; *Adv Exp Med Biol* 492 (2001) 11-17.

Bertini I., Cacciatore S., Jensen B.V., Schou J.V., Johansen J.S., Kruhoffer M., Luchinat C., Nielsen D.L., Turano P.; *Cancer Res*. 72 (2012) 356-364.

Bobe G., Murphy G., Albert P.S., Sansbury L.B., Lanza E., Schatzkin A., Cross A.J.; *Int J Cancer* 130 (2012) 1649-1659.

Brenner H., Bouvier A.M., Foschi R., Hackl M., Larsen I.K., Lemmens V., Mangone L., Francisci S., EUROCARE Working Group.; *Int J Cancer* 131 (2012) 1649-1658.

Bu Q., Huang Y.N., Yan G.Y., Cen X.B., Zhao Y.L.; *Comb Chem High T Scr* 15 (2012) 266-275.

O'Connell T.M.; *Bioanalysis* 4 (2012) 431-451.

Cai F., Dupertuis Y.M., Pichard C.; *Curr Opin Clin Nutr* 15 (2012) 99-106.

Chae Y.K., Kang W.Y., Kim S.H., Joo J.E., Han J.K., Hong B.W.; *Korean Chem Soc* 31 (2010) 379-383.

Chan E.C., Koh P.K., Mal M., Cheah P.Y., Eu K.W., Backshall A., Cavill R., Nicholson J. K., Keun H. C.; *J Proteome Res* 8 (2009) 352-361.

Chen G.C., Pang Z., Liu Q.F.; *Eur J Clin Nutr* 66 (2012) 1182-1186.

Cheng Y., Xie G., Chen T., Qiu Y., Zou X., Zheng M., Tan B., Feng B., Dong T., He P., Zhao L., Zhao A., Xu L. X., Zhang Y., Jia W.; *J Proteome Res* 3 (2012) 1354-1363.

Davis V.W., Bathe O.F., Schiller D.E., Slupsky C.M., Sawyer M.B.; *J Surg Oncol* 103 (2011) 451-459.

Denkert C., Budczies J., Weichert W., Wohlgemuth G., Scholz M., Kind T., Niesporek S., Noske A., Buckendahl A., Dietel M., Fiehn O.; *Mol Cancer* 18 (2008) 7-72.

Feng B., Yue F., Zheng M.H. Urinary markers in colorectal cancer, IN *Advances in Clinical Chemistry*. Makowski G. (Ed.) Academic Press, Farmington, (2009) 45-57.

Fernández-Arroyo S., Gómez-Martínez A., Rocamora-Reverte L., Quirantes-Piné R., Segura-Carretero A., Fernández-Gutiérrez A., Ferragut J.A.; *J Pharmaceut Biomed* 63 (2012) 128-134.

Fossel E.T., Carr J.M., McDonagh J.; *N Engl J Med* 315 (1986) 1369-1376.

Gerber M.; *Brit J Nutr* 107 (2012) S228-S23 9.

Gorham E.D., Garland C.F., Garland F.C., Grant W.B., Mohr S.B., Lipkin M., Newmark H.L., Giovannucci E., Wei M., Holick M.F.; *J Steroid Biochem* 97 (2005) 179-194.

Griffin J.L., Shockcor J.P.; *Nat Rev Cancer* 4 (2004) 551-561.

Hadziabdić M.O., Bozikov V., Pavić E., Romić Z.; *Coll Antropol* 36 (2012) 1427-1434.

Hansen L., Skeie G., Landberg R., Lund E., Palmqvist R., Johansson I., Dragsted L.O., Egeberg R., Johnsen N.F., Christensen J., Overvad K., Tjonneland A., Olsen A.; *Int J Cancer* 131 (2012) 469-478.

Herrero M., Plaza M., Cifuentes A., Ibáñez E.; *J Chromatogr A* 1217 (2010) 2512-2520.

Hirayama A., Kami K., Sugimoto M., Sugawara M., Toki N., Onozuka H., Kinoshita T., Saito N., Ochiai A., Tomita M., Esumi H., Soga T.; *Cancer Res* 1 (2009) 4918-4925.

Huxley R.R., Ansary-Moghaddam A., Clifton P., Czernichow S., Parr C.L., Woodward M.; *Int J Cancer* 125 (2009) 171-180.

Huxley R.R., Woodward M., Clifton P.; *Curr Nutr Rep* 2 (2013) 63-70.

Ibáñez C., Valdés A., García-Cañas V., Simó C., Celebier M., Rocamora-Reverte L., Gómez-Martínez A., Herrero M., Castro-Puyana M., Segura-Carretero A., Ibáñez E., Ferragut J.A., Cifuentes A.; *J Chromatogr A* 1248 (2012a) 139-153.

Ibáñez C., Simó C., García-Cañas V., Gómez-Martínez A., Ferragut J. A., Cifuentes A.; *Electrophoresis* 33 (2012b) 2328-2336.

Jiménez B., Mirnezami R., Kinross J., Cloarec O., Keun H.C., Holmes E., Goldin R.D., Ziprin P., Darzi A., Nicholson J.K.; *J Proteome Res* 12 (2013) 959-968.

Key T.J., Appleby P.N., Masset G., Brunner E.J., Cade J.E., Greenwood D.C., Stephen A.M.,

Kuh D., Bhaniani A., Powell N., Khaw K.T.; *Int J Cancer* 131 (2012) E320-E325.

Kim Y.S., Maruvada P., Milner J.A.; *Future Oncol* 4 (2008) 93-102.

Kondo Y., Nishiumi S., Shinohara M., Hatano N., Ikeda A., Yoshie T., Kobayashi T., Shiomi Y., Irino Y., Takenawa T., Azuma T., Yoshida M.; *Biomark Med* 5 (2011) 451-460.

Leichtle A.B., Nuoffer J.M., Ceglarek U., Kase J., Conrad T., Witzigmann H., Thiery J., Fiedler G.M.; *Metabolomics* 8 (2012), 643-653.

Li F., Qin X., Chen H., Qiu L., Guo Y., Liu H., Chen G., Song G., Wang X., Li F., Guo S., Wang B., Li Z.; *Rapid Commun Mass Spectrom* 27 (2013) 24-34.

Ludwig C., Ward D.G., Martin A., Viant M.R., Ismail T., Johnson P.J., Wakelam M.J., Günther U.L.; *Magn Reson Chem* 47 (2009) S68-S73.

Ma Y.L., Liu W.J., Peng J.Y., Zhang P., Chen H.Q., Qin H.L.; *Zhonghua Wei Chang Wai Ke Za Zhi* 12 (2009a) 386-390.

Ma Y.L., Qin H.L., Liu W.J., Peng J.Y., Huang L., Zhao X.P., Cheng Y.; *Dig Dis Sci* 54 (2009b) 2655-2662.

Ma Y.L., Liu W.J., Peng J.Y., Huang L., Zhang P., Zhao X.P., Chen Y., Qin H.L.; *Mol Biol Rep* 37 (2010) 1403-1411.

Maddula S., Baumbach J.J.; *Metabolomics* 7 (2011) 509-523.

Mal M., Koh P.K., Cheah P.Y., Chan E.C.Y.; *Anal Bioanal Chem* 403 (2012) 483-493.

Markuszewski M.J., Struck W., Waszczuk-Jankowska M., Kaliszan R.; *Electrophoresis* 31 (2010) 2300-2310.

Metallo C.M.; *Cancer Prev Res (Phila)* 5 (2012) 1337-1340.

Monleón D., Morales J.M., Barrasa A., López J.A., Vázquez C., Celda B.; *NMR Biomed* 22 (2009) 342-348.

Montrose D.C., Zhou X.K., Kopelovich L., Yantiss R.K., Karoly E.D., Subbaramaiah K., Dannenberg A.J.; *Cancer Prev Res (Phila)* 5 (2012) 1358-1367.

Mosby T.T., Cosgrove M., Sarkardei S., Platt K.L., Kaina B.; *Anticancer Res* 32 (2012) 4171-4192.

Nagrath D., Caneba C., Karedath T., Bellance N.; *BBA-Bioenergetics* 1807 (2011) 650-663.

Nambiar P.R., Gupta R.R., Misra V.; *Mutat Res* 693 (2010) 3-18.

Nishiumi S., Kobayashi T., Ikeda A., Tomoo Y., Megumi K., Yoshihiro I., Tatsuya O., Nobuhide H., Seiji K., Tadaomi T., Takeshi A., Masaru Y.; *PLoS ONE* 7 (2012) e40459.

Ng D.J.Y., Pasikanti K.K., Chan E.C.Y.; *Metabolomics* 7 (2011) 155-178.

Phua L.C., Mal M., Koh P.K., Cheah P.Y., Chan E.C., Ho H.K.; *Cancer Chemother Pharmacol* 71 (2013) 817-823.

Qiu Y., Cai G., Su M., Chen T., Liu Y., Xu Y., Ni Y., Zhao A., Cai S., Xu L.X., Jia W.; *J Proteome Res* 9 (2010) 1627-1634.

Rao A., Janezic S.; *Nutr Cancer* 18 (1992) 43-52.

Ritchie S., Ahiahonu P.W.K., Jayasinghe D., Heath D., Liu J., Lu, Y., Jin W., Kavianpour A., Yamazaki Y., Khan A.M., Hossain M., Su-Myat K.K., Wood P.L., Krenitsky K., Takemasa I., Miyake M., Sekimoto M., Monden M., Matsubara H., Nomura F., Goodenowe D.B.; *BMC Med* 8 (2010) 13.

Robertson D.G., Watkins P.B., Reily M.D.; *Toxicol Sci* 120 (2011) S146–S170.

Rudolf E., Andelova H., Cervinka M.; *Anti-Cancer Agent Me* 7 (2007) 559-575.

Sant M., Allemani C., Sieri S., Krogh V., Menard S., Tagliabue E., Nardini E., Micheli A., Crosignani P., Muti P., Berrino F.; *Int J Cancer* 121 (2007) 911-914.

Serkova N.J., Glunde K.; *Methods Mol Biol* 520 (2009) 273-295.

Simó C., Ibáñez C., Gómez-Martínez A., Ferragut J.A., Cifuentes A.; *Electrophoresis* 32 (2011) 1765-1777.

Stewart B.W., Kleihues P., (Eds.) *World cancer report*. IARC Press, Lyon (2003).

Tavani A., Malerba S., Pelucchi C., Dal Maso L., Zucchetto A., Serraino D., Levi F., Montella M., Franceschi S., Zambon A., La Vecchia C.; *Ann Oncol* 23 (2012) 2737-2742.

Tessem M.B., Selnaes K.M., Sjursen W., Trano G., Giskeodegard G.F., Bathen T.F., Gribbestad I.S., Hofsli E.; *J Proteome Res* 9 (2010) 3664-3670.

Tetè S., Nicoletti M., Saggini A., Maccauro G., Rosati M., Conti F., Cianchetti E., Tripodi D., Toniato E., Fulcheri M., Salini V., Caraffa A., Antinolfi P., Frydas S., Pandolfi F., Conti P., Potalivo G., Theoharides T.C.; *Int J Immunopathol Pharmacol* 25 (2012) 573-581.

Thomson C.A.; *Nutr Clin Pract* 27 (2012) 636-650.

Valdés A., García-Cañas V., Rocamora-Reverte L., Gómez-Martínez A., Ferragut J.A., Cifuentes A.; *Genes Nutr* 8 (2013) 43-60.

Verma M., Khoury M.J., Ioannidis J.P.; *Cancer Epidemiol Biomarkers Prev* 22 (2013) 189-200.

Wang H., Tso V.K., Slupsky C.M., Fedorak R.N.; *Future Oncol* 6 (2010) 1395-1406.

Wang H., Schiller D.E., Tso V.K., Slupsky C.M., Wong C.K., Fedorak R.N.; *Gastroenterology* 140 (2011) S-40.

Wang Z.J., Joshi A.M., Ohnaka K., Morita M., Toyomura K., Kono S., Ueki T., Tanaka M., Kakeji Y., Maehara Y., Okamura T., Ikejiri K., Futami K., Maekawa T., Yasunami Y., Takenaka K., Ichimiya H., Terasaka R.; *Nutr Cancer* 64 (2012) 798-805.

Wark P.A., Lau R., Norat T., Kampman E.; *Am J Clin Nutr* 96 (2012) 622-631.

Williams C.D., Satia J.A., Adair L.S., Stevens J., Galanko J., Keku T.O., Sandler R.S.; *Nutr Cancer* 62 (2010) 701-709.

Williams M.D., Reeves R., Resar L.S., Hill H.H.Jr.; *Anal Bioanal Chem* (2013) (doi:10.1007/s00216-013-6777-5). In press.

Wu S.J., Feng B., Li K., Zhu X., Liang S.H., Liu X.F., Han S., Wang B.L., Wu K.C., Miao D.M., Liang J., Fan D.M.; *Am J Med* 125 (2012) 551-559.

Yoshie T., Nishiumi S., Izumi Y., Sakai A., Inoue J., Azuma T., Yoshida M.; *Cancer Sci* 103 (2012) 1010-1021.

Zimmermann D., Hartmann M., Moyer M.P., Nolte J., Baumbach J.I.; *Metabolomics* 3 (2007) 13-17.

3.2. **CAPÍTULO 1 (Chapter 1)**

Is metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles

Carolina Simó[1]
Clara Ibáñez[1]
Ángeles Gómez-Martínez[2]
José A. Ferragut[2]
Alejandro Cifuentes[1]

[1]Institute of Food Science Research (CIAL), CSIC, Nicolas Cabrera, Campus de Cantoblanco, Madrid, Spain
[2]Institute of Molecular and Cellular Biology, Miguel Hernandez University, Avda, Universidad s/n; Elche, Alicante, Spain

Research Article

Is metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles

In this work, four different metabolite purification approaches are investigated prior to metabolomics of human HT29 colon cancer cells. Namely, methanol deproteinization, ultrafiltration and two SPE methods using C18 and polymer-based cartridges were studied. The extracts were characterized via a metabolomic approach based on the application of CE TOF MS (CE-MS). CE-MS analysis time was less than 20 min per sample and allowed the simultaneous and reproducible analysis of more than 80 metabolites in a single run with a minimum consumption of sample and reagents. Metabolome analysis revealed in some cases important differences among the studied metabolite purification procedures. No significant differences were observed in the metabolite profile using C18 and polymer-based cartridges, or between ultrafiltration and methanol deproteinization. However, important differences were observed in the metabolomic profiles obtained from SPE and methanol deproteinization samples. These results demonstrate the crucial role of the metabolite purification strategy in metabolomics since it can bias (and in some cases mislead) the conclusions achieved by the metabolomic study.

1 Introduction

The general aim of any metabolomic study applied to a living cell, tissue or biofluid is to gain detailed insight of the molecular mechanisms governing the metabolic pathways. Like proteome, metabolome is dynamic and highly variable among cell types, organisms, individuals, environments etc. Thus, one of the main challenges in metabolomics is to overcome the chemical complexity, heterogeneity and wide dynamic concentration range of endogenous metabolites (amino acids, amines, small peptides, nucleic acids, carbohydrates, organic acids, vitamins, steroids, coenzymes etc.) present in a biological sample. Metabolomics presents also unique challenges for separation and detection techniques. No single analytical methodology or platform is available to detect, quantify and identify all metabolites in a certain sample. Two analytical platforms are currently used for metabolomic analyses: MS and NMR-based systems. NMR, that in some cases does not require previous analyte separation (and requires minimal sample precleaning), provides detailed information on the molecular structure of compounds complementary to MS-based metabolomic data, although at the expense of low sensitivity. Improved mass spectrometers with better sensitivity and superior mass accuracy and resolution aim at the identification and quantitation of complex metabolite mixtures in a single experiment. The use of high and ultra-high resolution mass analyzers (TOF, FTMS, Orbitrap® etc) is essential to obtain accurate mass measurements for the determination of elemental compositions of metabolites and to carry out tentative identification based on metabolites databases. On the other hand, MS/MS or MS^n experiments, especially when product ions are accurately analyzed at high resolution (namely, Q-TOF, TOF-TOF, LTQ-Orbitrap®), provide useful additional structural information for the

Correspondence: Professor Alejandro Cifuentes, Institute of Food Science Research (CIAL), CSIC, Nicolas Cabrera 9, Campus de Cantoblanco, 28049 Madrid, Spain
E-mail: a.cifuentes@csic.es
Fax: +34-910017905

Abbreviations: ABN, acidic, basic and neutral; **AEBSF,** 4-(2-aminoethyl)benzenesulfonyl fluoride; **EIE,** extracted ion electropherogram; **GABA,** γ-aminobutyric acid; **KEGG,** Kyoto Encyclopedia of Genes and Genomes

identification of the metabolites. These techniques either stand-alone or combined with separation techniques (typically, LC-NMR, GC-MS, LC-MS and CE-MS) are capable of producing complementary analytical information to get a more extensive metabolome coverage [1–5]. Systematic profiling/fingerprinting of as many metabolites as possible has gained broad interest during the last decade. CE is particularly suited for the rapid separation of ionic and polar compounds with very high resolution using extremely small reagents and sample volumes. Moreover, no pre-column derivatization of analytes is necessary. On the contrary, lower sensitivity and higher variability of migration times are generally obtained compared to LC or GC. TOF mass analyzer is preferably used in CE-MS due to its high spectral acquisition rate and high mass resolution, which perfectly fits with the narrow peaks provided by CE. The capacity of CE-MS to analyze complex mixtures of metabolites in short times opens interesting possibilities in the growing Metabolomic area. Until today, CE-MS has successfully been applied in many metabolomic studies. Interesting reviews have already been published describing the potential of CE-MS in metabolomics [6–9].

In general, non-targeted metabolomics is addressed to detect as many metabolites as possible in a certain sample. However, at present there is not a general applicable sample preparation protocol to extract the whole range of endogenous metabolites present in a biological sample. Sample preparation is a critical step in any analytical method with important consequences in the final results. In order to prevent loss of metabolites in non-targeted metabolomics minimal sample treatment should be carried out before analysis. The procedure used for metabolite extraction has to be robust and highly reproducible. It will depend on both the sample type and the targeted metabolites of interest (fingerprinting or profiling approach). Most sample extraction techniques are more or less selective; thus, the choice of the appropriate one is very critical for metabolomic studies since it can bias the final results obtained. In a non-targeted metabolomic work there is a clear need for the development of methods that enable a comprehensive characterization of the metabolome. In this work, CE-MS was used to study the metabolome of human HT29 colon cancer cells. For non-targeted metabolomics of biological samples deproteinization with an organic solvent is often carried out to avoid adsorption of proteins to the inner capillary wall and the consequent damage on metabolite separation and ion suppression at the ion source. In order to minimize sample handling and avoid extra variability to the final statistic study required in any non-targeted metabolomic study, several common extraction procedures (protein precipitation, SPE, filtration) were selected among the enormity of possibilities due to their simplicity, effectiveness and recognized reproducibility. Namely, four metabolite purification approaches were systematically compared in this work (i.e. two different SPE methods, protein precipitation with methanol and ultrafiltration). The metabolomic profiles obtained were compared based on the total number and type of extracted

metabolites, using the information provided by CE-MS for their tentative identification. The results give an additional proof on the difficulty to achieve a representative metabolite profile in metabolomics.

2 Materials and methods

2.1 Reagents

All chemicals were of analytical reagent grade and used as received. Methanol used in the metabolite extraction procedure was from Sigma-Aldrich (St. Louis, MO). All reagents and solvents employed in the preparation of CE electrolytes and sheath liquids were of MS grade: formic acid and 2-propanol were from Riedel-de Haën (Seelze, Germany), and water was from Scharlau (Barcelona, Spain). Amino acids from Sigma-Aldrich were dissolved in purified water deionized by using a Milli-Q system from Millipore (Bedford, MA, USA), at the following concentrations: 0.58 mg/mL arginine, 0.49 mg/mL lysine, 0.52 mg/mL histidine, 0.34 mg/mL γ-aminobutyric acid (GABA), 0.39 mg/mL valine, 0.35 mg/mL serine, 0.44 mg/mL leucine, 0.40 mg/mL threonine, 0.49 mg/mL glutamine, 0.38 mg/mL proline, 0.44 mg/mL aspartic acid and 0.09 mg/mL tyrosine. Tyramine, DL-methionine sulfone and 1,4-piperazinediethanesulfonic acid from Sigma-Aldrich were selected as internal standards.

2.2 Samples

Human colorectal adenocarcinoma HT29 cells were used in all the experiments. HT29 cells were grown in DMEM supplemented with 5% heat-inactivated fetal calf serum, 2 mM of L-glutamine, 50 U/mL of penicillin G and 50 μg/mL of streptomycin, at 37°C in humidified atmosphere and 5% CO_2.

For sample preparation, a PBS solution containing 138 mM sodium chloride, 2.7 mM potassium chloride and 10 mM sodium hydrogen phosphate, at pH 7.4, was purchased from Sigma-Aldrich. Composition of homogenization buffer was next: 10 mM Tris-HCl, 5 mM EDTA, 120 mM NaCl, at pH 7.4, all of them from Sigma-Aldrich. A protease inhibitor cocktail containing 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF), pepstatin A, E-64, bestatin, leupeptin and aprotinin were purchased from Sigma-Aldrich.

Human HT29 colon cancer cells were washed with PBS solution and centrifuged. The pellet was resuspended with homogenization buffer and protease inhibitor cocktail. Cells were disrupted with a Polytron homogenizer and centrifuged (14 min at 14 000 g and 4°C).

Pellet (nuclear fraction) was discarded and supernatant was centrifuged for 1 h at 100 000 g and 4°C. Supernatant (cytosolic fraction) was stored at −80°C until metabolite purification procedure was carried out. At this temperature, enzyme activity is stopped and samples can safely be stored without continuing metabolic activity. The total protein

content was determined by the Bradford method using a commercial dye reagent from Bio-Rad (Hercules, CA, USA) and using BSA as standard.

2.3 Metabolite extraction procedures

Four different metabolite extraction procedures were studied in this work: two SPE methods using different sorbents, protein precipitation with methanol and ultrafiltration.

2.3.1 SPE

Two different sorbents were investigated, namely, Isolute C18 Endcapped cartridges (100 mg) from Biotage (Cardiff, Wales, UK), and Evolute™ acidic, basic and neutral (ABN) columns (25 mg). Lower sorbent mass was selected for the last as a result of the higher capacity of the polymeric sorbents due to their higher specific surface area. In both cases, activation, conditioning and elution were carried out in identical conditions. Activation and conditioning were carried out by passing 1 mL of methanol followed by 1 mL of water through the cartridges. About 350 µL of water were added to 150 µL of cytosolic fraction obtained from the cell culture. This solution was loaded onto the column. After sample loading sorbent was washed with 1 mL of water-methanol (95:5 v/v). Sample elution was performed with 500 µL of methanol. Eluted sample was then aliquot in 100 µL volume.

2.3.2 Protein precipitation

For protein precipitation, 350 µL of methanol were added to 150 µL of cytosolic fraction sample obtained from the cell culture. The solution was incubated at −20°C for 2 h. After incubation the suspension was centrifuged at 20000 g and 4°C for 5 min. The pellet was discarded and the supernatant fraction was collected and aliquot in 100 µL volume.

2.3.3 Ultracentrifugation

About 350 µL of water were added to 150 µL of cytosolic fraction sample obtained from the cell culture. This solution was ultrafiltrated using an Amicon Ultra 3 kDa centrifugal device (70 min at 14 000 g and 4°C) from Millipore. Fraction with molecular weight lower than 3 kDa was collected and aliquot in 100 µL volume.

In all cases, after metabolite purification procedures, the obtained 100 µL aliquots were vacuum-dried. The dried extracts were stored at −80°C until used. Prior to CE-MS analysis, dried extracts were dissolved in 20 µL of water of MS grade from Scharlau.

2.4 Instrumentation

CE analyses were carried out in a P/ACE 5500 CE apparatus from Beckman Instruments (Fullerton, CA, USA). The instrument was controlled by a PC running the System Gold software from Beckman. Uncoated fused-silica capillaries (50 µm id and 90 cm total length) from Composite Metal Services (Worcester, England) were coupled to MS through an orthogonal ESI interface model G1607A from Agilent Technologies (Palo Alto, CA, USA). Electrical contact at the electrospray needle tip was established via a sheath liquid. A TOF MS instrument (micrOTOF) from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a PC running the micrOTOF control software from Bruker Daltonics.

2.5 CE-ESI-TOF MS conditions

Before first use, the separation capillary was conditioned by rinsing with 1 NaOH for 10 min, followed by 20 min with Milli-Q water and 5 min with the separation buffer. After each run, the capillary was conditioned with Milli-Q water for 4 min, followed by separation buffer for 4 min. Injections (53 nL of sample volume) were made at the anodic end using N_2 pressure of 0.5 psi (34.5 mbar) for 80 s. The electrophoretic separation was achieved using +25 kV as running voltage at a constant temperature of 25°C in a 1 M formic acid BGE. Electrical contact at the electrospray needle tip was established via a sheath liquid based on isopropanol-water (50:50 v/v) and delivered at a flow rate of 0.24 mL/min by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). The mass spectrometer operated with the ESI source in the positive ion mode. The nebulizer and drying gas conditions were 0.4 bar N_2 and 4 L/min N_2, respectively, and maintaining the ESI chamber at 200°C. The micrOTOF was operated to acquire spectra in the m/z range of 50–600 every 90 ms. The accurate mass data of the molecular ions were processed using the DataAnalysis 3.3 software from Bruker Daltonics. External and internal calibration of the TOF MS instrument was performed by introducing a 5 mM sodium formate solution through the separation capillary. Masses for the calibration of the TOF MS instrument were next: 90.9766, 158.9641, 226.9515, 294.9389, 362.9263, 430.9138, 498.9012 and 566.8886 m/z. Each sample was analyzed in triplicate by CE-MS.

2.6 Data processing

Calculation of the elemental composition of compounds was carried out using the Generate Molecular Formula Editor within DataAnalysis software from Bruker Daltonics. Accurate m/z value and migration time from each peak were annotated. Redundant responses from the same ion, such as isotopic peaks, fragments, adducts, dimers etc, were removed based on established m/z differences. Spike noise and low reliability signals (no peak-like shape) were also eliminated. For the calculation of the total number of the different metabolites detected by CE-MS after each extraction protocol, only those metabolites that repeatedly

appeared in three consecutive replicates were considered. TOF MS provided a high mass resolution and high mass accuracy with errors usually below 5–10 ppm. Selected mass spectra were processed through the software DataAnalysis, which provided a list of possible elemental formulas by using the Generate Molecular Formula Editor, which provided standard functionalities such as minimum/ maximum elemental range, electron configuration and ring-plus double bonds equivalents, as well as a comparison between the theoretical and the experimental isotopic pattern (Sigma-Value™) for increased confidence in the theoretical molecular formula. Tentative identification based on the obtained theoretical molecular formula was carried out with different free available databases: Human Metabolome Database (HMDB) [10], Metlin [11], Kyto Encyclopedia of Genes and Genomes (KEGG) compound [12–14], and PubChem (http://pubchem.nc-bi.nlm.nih.gov/). In cases where databases offered more than one possible metabolite for one molecular formula, migration time provided by electrophoretic separation was studied to elucidate the expected electrophoretic mobility of that compound at the separation pH. When available, standards were used to confirm metabolite identification. Metabolite relative levels were calculated from CE-MS data using peak areas. Freely MassTRIX web server [15] was used to map the identified metabolites to KEGG pathways.

3 Results and discussion

3.1 Sample preparation

Considering the great diversity and heterogeneity of metabolites, simultaneous purification and analysis of all metabolites from a biological sample is still a challenge. Sample preparation (extraction, clean-up, concentration etc.) for metabolomic studies depends on the type of sample being analyzed, the subsequent method of analysis and the goal of the metabolic work (whether targeted or non-targeted metabolomic study is going to be carried out). In many cases, sample must be pre-treated in order to eliminate interfering matrix constituents, which can negatively affect the metabolomic results.

In CE-MS, complex biological samples containing components that interact with the inner silica capillary surface lead to alterations in EOF and subsequent lack of reproducibility between injections. More precisely, one of the main problems when working with biological samples is the adsorption of proteins onto the capillary wall that may produce changes on EOF. On the other hand, high-salt-containing sample could reduce efficiency due to electro-migration dispersion, producing irreproducible injections (in case of electrokinetic injection) and migration times. Matrix effects can also lead to difficulties in the detection of certain compounds by MS since ESI-MS detection is sensitive to the presence, e.g. of salts or other compounds that can comigrate with the analytes of interest.

In this work, a lysate from human HT29 colon cancer cells with rather high salt content (see Section 2.2) and a total protein concentration of 0.9 mg/mL was analyzed using different sample preparation methodologies. Two different procedures were assayed to remove proteins from these samples: protein precipitation with methanol and deproteinization by ultrafiltration using a membrane pore size of 3 kDa. Using any of the mentioned procedures, only deproteinization is carried out, while salts remained in the final extract. On the other hand, offline SPE was also assayed for both sample deproteinization and desalting of samples, while additionally it can enrich the more retained analytes. A wide range of chemically modified sorbent materials (silica gel or synthetic resins, modified or not) enable metabolite purification based on different types of physicochemical interactions. Reversed-phase was selected as it is widely used in many applications including meta-bolomics. The use of polymeric sorbents is gaining more attention as a method for metabolite purification in non-targeted approaches. Thus, two different SPE sorbents were studied in this work: a classical octadecyl bonded endcapped silica (C18) sorbent and a polymer-based sorbent (ABN). It is expected that the C18 SPE retains compounds of mid to low polarity due to their polar interaction with the sorbent. On the other hand, ABN uses a polystyrene-divinylbenzene sorbent, suitable for the extraction of a wide range of analytes (ABN). The same extraction protocol was selected for both types of sorbents. In SPE loss of metabolites during sample loading and washing steps is generally unavoidable. In our case, washing fractions can contain small hydrophilic and charged metabolites that will be lost. However, together with these potentially interesting metabolites, interfering compounds as salts and proteins from the cytoplasm are expected to be also present. We decided to discard this fraction in order to avoid protein adsorption onto the inner capillary wall, which could affect migration time reproducibility and metabolite signal suppression at the ESI ion source. The same cytosol sample was subjected to these four purification procedures. After metabolite purification, extracts were directly analyzed by CE-MS.

3.2 CE-ESI-TOF MS method development

A low pH was selected for CE-MS analysis in order to both avoid analyte adsorption onto the inner capillary wall and confer positive charge to the analytes to improve their ESI ionization yield. A BGE composed of 1 M formic acid in water at pH 1.8 was used. At the low pH electrolyte used in this work most amino acids, amines, nucleosides, small peptides, and in general, basic compounds, present overall positive charge and they migrate before EOF. Using these analytical conditions, CE-MS method is focused on cationic metabolites. However, using these electrophoretic conditions it was also possible to detect some acidic compounds (bearing simultaneously negative and positive charge in

their structure), as they were carried to the MS by the residual EOF.

To overcome any influence from the different salt concentrations of the purified samples on the amount injected, pressure injection was applied instead of electrokinetic injection, in all the cases the injected sample plug being 2.7 cm (3% of the capillary length). We have also taken into account that due to the complexity of the sample, during the CE-MS analysis it is possible that some metabolites can comigrate, which disturbs the ionization conditions since the presence of strongly ionizable compounds will suppress the signal of the less ionizable ones. On the other hand, a disadvantage of CE-MS methods compared to LC-MS or GC-MS methods is the lower repeatability of peak areas. Nevertheless, good repeatability was obtained using this CE-MS method with %RSD values lower than 12% in the intra-day ($n = 5$ injections) and inter-day (3 days, $n = 15$ injections) repeatability study of ten different metabolites arbitrarily selected from the CE-MS electropherogram (peak area values and migration times are given as Supporting Information in Table S1).

3.3 CE-ESI-TOF MS metabolic analysis of human HT29 colon cancer cells

3.3.1 Matrix effect

After metabolite purification, the four different extracts from the HT29 colon cancer cells were analyzed by CE-MS. Although injected sample volume was the same in all the cases (3% of the capillary length), sample conductivity was expected to be different depending on the sample matrix due to their different salt and metabolite content. In order to study this effect on the CE-MS metabolite profile, the following experiments were carried out. First, formation of sodium formate clusters during CE-MS analysis of desalted (ABN extract) and non-desalted (MeOH extract) samples was observed (extracted ion electropherograms, EIEs, are given in Fig. S1 as Supporting Information). Basically, the results showed a small narrow peak in the desalted sample, while a broad band migrating from minutes 4 to 7 was observed in the non-desalted sample. It was also confirmed that the width of this band was proportional to the injected sample volume (data not shown). Typical sodium formate clusters observed were Na(NaCOOH)1 (90.9766 *m/z*), Na(NaCOOH)2 (158.9641 *m/z*), Na(NaCOOH)3 (226.9515 *m/z*), Na(NaCOOH)4 (294.9389 *m/z*), Na(NaCOOH)5 (362.9263 *m/z*), Na(NaCOOH)6 (430.9138 *m/z*), Na (NaCOOH)7 (498.9012 *m/z*) and Na(NaCOOH)8 (566.8886 *m/z*), which are produced by the interaction of the sodium ion from the sample with the formate counterions from the BGE. This clusters formation during the CE separation is also expected to have some influence on the CE-MS metabolite profile obtained for these samples. In order to study this effect, EIEs of common metabolites (identical *m/z*) to all extracts were represented (results are given as

Supporting Information in Fig. S2, indicating with letters from A to G the main obtained peaks). As expected, a migration time shift among the different extracts (ABN, C18, MeOH and ultrafiltration) was obtained. This effect is dependent on the ionic strength of the sample, which leads to different electrical conductivities and, as a result, different effective electric fields inside the capillary. In addition to the observed migration time shift, a slight increase in plate number of the peaks for the high-ionic-strength samples (non-desalted MeOH and ultrafiltration extracts) was obtained. This effect is contrary to the expected from theory (higher electromigration dispersion in the more salty samples) and although at the moment we do not see a clear reason, it could be explained through a stacking process induced by the sodium clusters band mentioned above [16].

3.3.2 Comparison of metabolite extraction methods

Base peak electropherograms from the four different metabolite extracts were obtained by CE-MS (they are shown as Supporting Information in Fig. S3). Comparing the two extracts obtained from SPE protocols (using ABN and C18 sorbents) it was observed that both CE-MS profiles were very similar. Namely, 80 compounds were detected by CE-MS in the ABN extract and 71 compounds in the C18 extract, observing that 62 of these compounds were common to both extracts. It is interesting to highlight at this point the better metabolite extraction power of the ABN sorbent at the selected SPE equilibration/elution conditions based on the higher number of metabolites detected (80) compared to the number obtained after C18 extraction (71).

Regarding the comparison of the metabolite extraction when only sample deproteinization was carried out (MeOH protein precipitation versus ultrafiltration), the total number of compounds identified were 83 in the MeOH extract and 74 after ultrafiltration. Among all these compounds, 72 were common species. This result demonstrates the similarity of both deproteinization procedures. It is important to remark that sample treatment using protein precipitation with MeOH is simpler and less expensive protocol than ultrafiltration.

Next, selectivity of metabolite purification was studied comparing the best two protocols (which allowed the detection of a higher number of metabolites), namely, deproteinization and desalting using ABN cartridge and MeOH deproteinization. After mass spectra analysis, important differences in terms of the selectivity of metabolite extraction were observed. In Fig. 1, the EIEs of the compounds with different migration times observed in both extracts (ABN and methanolic extract) are represented. Taking into account both extracts, a total number of 115 different metabolites were detected, of which 80 metabolites were found in ABN, 83 in MeOH extracts, 48 compounds being common to both extracts. In Fig. 2, a bar plot of peak areas of the common metabolites extracted using the ABN cartridge or MeOH protein precipitation is presented. On
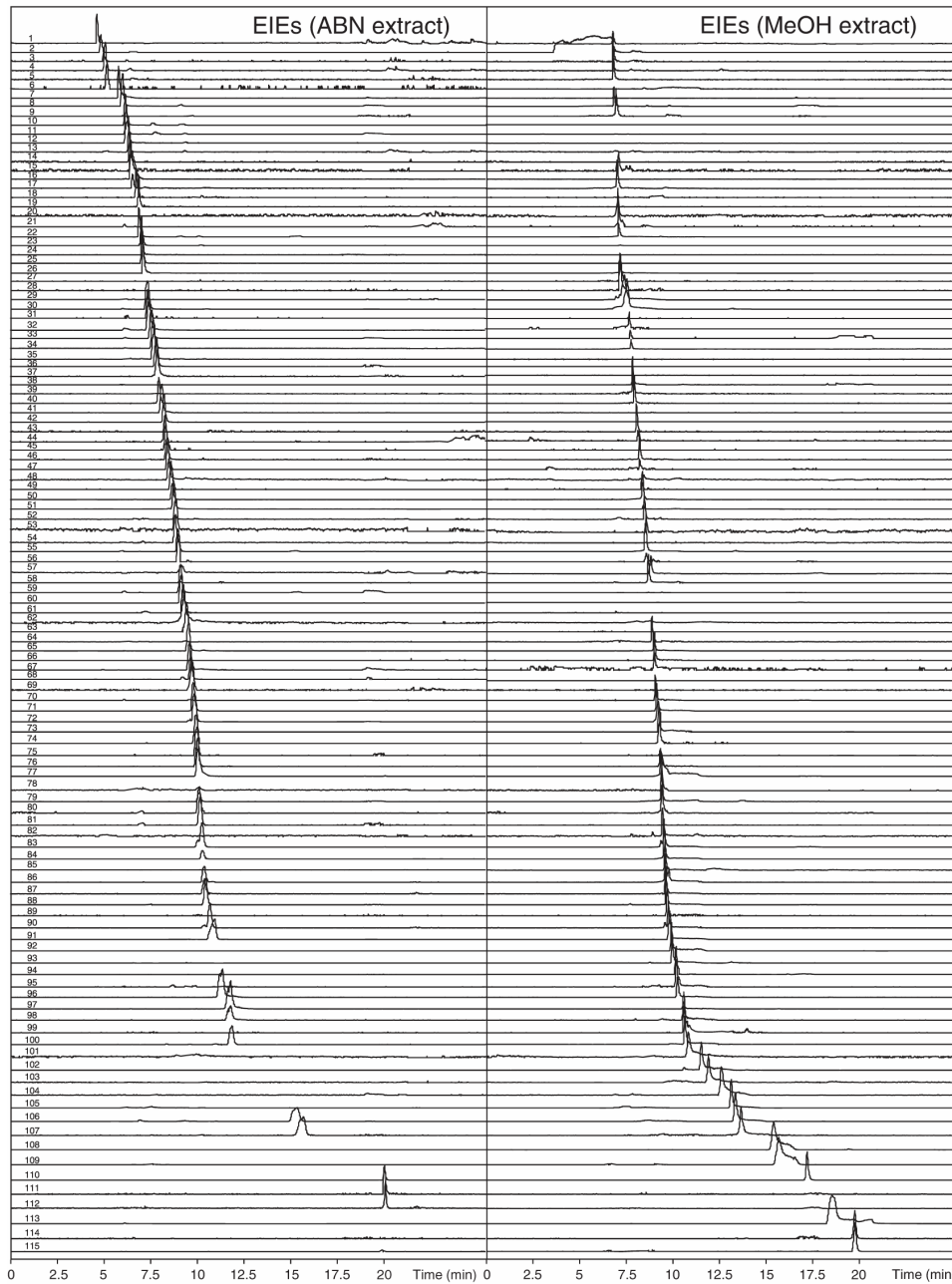
Figure 1. CE-MS EIEs of the 115 *m/z* values detected in metabolite extracts obtained from SPE (ABN cartridge) or methanol deproteinization. CE-MS conditions are described in the text.

the top of each column the metabolite (or peak) number assigned in Fig. 1 and Table 1 is indicated. From this group of 48 compounds, statistically significant higher intensities (after ANOVA at 5% significance level) were observed for 22 peaks (namely, 1, 8, 22, 26, 33, 34, 50, 51, 53, 56, 65, 67, 72, 76, 77, 83, 86, 88, 90, 91, 96 and 97) when ABN extraction was used. On the other hand, among the common 48 compounds, only eight showed a significantly higher

$(p < 0.05)$ signal in the MeOH extract (peaks 17, 30, 52, 55, 57, 79, 84 and 110), which seems to indicate some clear bias in the quantitative extraction of these compounds depending on the purification protocol applied. Regarding the non-common compounds, 32 metabolites were detected using SPE with the ABN column, and 35 metabolites could only be detected by sample deproteinization with MeOH. This result is indicative of an important bias in the quantitative

Figure 2. Representation of bar plot of peak areas of the common metabolites extracted using SPE (ABN cartridge) or protein precipitation with methanol. Metabolite number is assigned in Fig. 1 and Table 1.
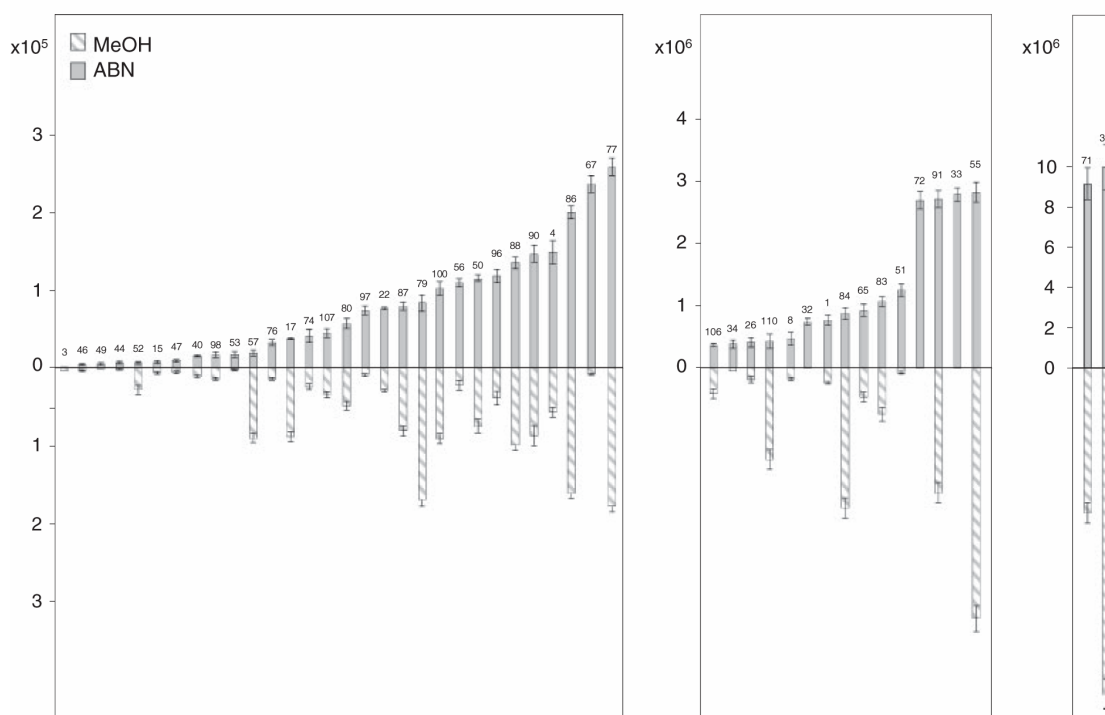
metabolomic analysis induced by the sample preparation step, as well as in the type and nature of the identified metabolites as will be corroborated below. The reproducibility obtained after metabolite purification using ABN or MeOH was similar. Namely, RSD values for peak areas were lower than 12 and 13% after ABN and MeOH extraction, respectively. These values were calculated for metabolites 56, 72 and 88 in three different extracts obtained under the same conditions using ABN or MeOH purification and analyzed in triplicate in the same day (total $n = 9$). This information is given as Supporting Information in Table S2.

3.3.3 Metabolite identification

In Table 1, a list of detected compounds and tentative identification using both ABN and MeOH purification protocols is presented. The TOF-MS analyzer used in this work was operated to acquire spectra in the range of 50–600 m/z. Most of the observed ions were single protonated ions $(M+H^+)$ and few of the species were detected as doubly charged (see Table 1). A total number of 115 different metabolites were detected with molecular masses from 90.0 to 777.2 Da. Among them, 45 compounds could not be identified since each one of the obtained m/z values was associated with more than ten molecular formulas (considering an error lower than 10 ppm). In other cases a molecular formula was proposed but no endogenous

metabolite could be found in metabolite databases. In certain cases, two or three metabolites were associated to a single molecular formula (metabolites 21, 29, 32, 38, 39, 52, 57, 64, 66, 70, 85, 108). In these cases, confirmation of identification was carried out based on their expected electrophoretic mobility at the separation pH and/or, when available, using standards. Thus, identification of histidine (peak 15), arginine (peak 20), lysine (peak 21), GABA (peak 29), valine (peak 52), serine (peak 53), leucine (or isoleucine) (peak 57), threonine (peak 64), glutamine (peak 66), proline (peak 70), tyrosine (peak 78) and aspartic acid (peak 85) was corroborated mixing the corresponding metabolite extract with standards solution. In certain cases, expected electrophoretic mobility was not enough information, and standard compound was not available in our laboratory. Thus, more than one metabolite had to be associated to a unique molecular formula. As an example, peak 32 with m/z value of 152,1072 and formula $C_9H_{13}NO$ was tentatively identified as *N*-methyltyramine and *N*-methylphenylethanolamine. Similar situation was founded for peaks 29, 38, 39 and 108 (see Table 1).

It is interesting to mention that among all amino acids detected (12), only histidine, valine, serine and leucine (or isoleucine) were observed in both metabolite extracts obtained after MeOH deproteinization or ABN purification. A more hydrophobic amino acid derivative as lipoillysine was only found in the ABN extract, which seems to indicate

Table 1. Tentative identification of metabolites after CE-MS analysis

No.	ABN	MeOH	m/z	z	Sigma™	Error (ppm)	Formula	Tentative identification	Classification	Database	KEGG pathway
1	Yes	Yes	145.0167	1	0.156	−2.9	$C_5H_5ClN_2O$	Disodium formate	Inorganic salt		
2	Yes	Yes	90.9768	1	0.001	0.2	CNa_2O_2				
3	Yes	Yes	272.9392	1	<0.1	<10	—[a]				
4	Yes	Yes	189.0972	1	0.102	2.4	$C_6H_{12}N_4O_3$	X[b]			
5	No	Yes	225.1431	1		+10					
6	Yes	No	427.2860	1	<0.1	<10	—				
7	Yes	No	134.9864	1	0.025	−4.1	$C_2H_2N_2O_3S$	X			
8	Yes	Yes	321.1257	1	0.026	−9.2	$C_{11}H_{20}N_4O_5S_1$	Tripeptide (G,N,M; A,Q,C)	Polypeptide	METLIN	
9	No	Yes	116.0837	1		+10					
10	Yes	No	335.1442	1	0.026	4.6	$C_{14}H_{26}N_2O_3S_2$	Lipoyllysine	Lipoic acid derivative	HMDB12996	
11	Yes	No	363.1763	1	0.04	3.6	$C_{16}H_{22}N_6O_4$	Thyrotropin-releasing factor	Polypeptide	HMDB05763	hsa04080
12	Yes	No	175.0855	2	0.020	4.6	$C_{12}H_{24}N_6O_4S_1$	Tripeptide (C,R,A)	Polypeptide	METLIN	
13	Yes	No	377.1924	1	0.024	9.1	$C_{21}H_{28}O_6$	18-Oxocortisol	Steroid derivative	HMDB00332	
14	Yes	No	203.1115	2	0.027	−6.2	$C_{20}H_{28}N_4O_5$	Tripeptide (I/L, W, S; W, T, V)	Polypeptide	METLIN	
15	Yes	Yes	156.0767	1	0.175	0.7	$C_6H_9N_3O_2$	Histidine	Amino acid	HMDB00177[c]	hsa00340, hsa00410, hsa02010, hsa00970,
16	Yes	No	211.0943	1	<0.1	<10	—				
17	Yes	Yes	128.0818	1	0.006	−0.5	$C_5H_9N_3O$	X			
18	Yes	No	470.1381	1	<0.1	<10	—				
19	Yes	No	242.5777	2	<0.1	<10	—				
20	No	Yes	175.1193	1	0.046	−1.9	$C_6H_{14}N_4O_2$	Arginine	Amino acid	HMDB03416[c]	hsa00472, hsa00330, hsa02010, hsa00970
21	No	Yes	147.1133	1	0.042	−3.5	$C_6H_{14}N_2O_2$	Lysine (+2 metabolites)	Amino acid	HMDB00182[c]	hsa00780, hsa00310, hsa00300, hsa02010, hsa00970
22	Yes	Yes	504.1611	1	<0.1	<10	—				
23	Yes	No	546.2033	1	0.088	−0.9	$C_{20}H_{35}NO_{16}$	Lacto-*N*-triaose	Carbohydrate	HMDB06592	
24	Yes	No	194.0089	1	<0.1	<10	—				
25	Yes	No	532.1955	1	<0.1	<10	—				
26	Yes	Yes	160.1078	1	0.003	1.6	$C_6H_{13}N_3O_2$	X			
27	No	No	526.1438	1	<0.1	<10	—				
28	No	No	510.1496	1	<0.1	<10	—				
I.S.	Yes	Yes	138.0917	1	0.004	−2.9	$C_8H_{11}NO$	Tyramine	Amino acid	HMDB00112[c]	
29	No	No	104.0704	1	0.005	2.2	$C_4H_9NO_2$	Aminobutyric acid (+7 metabolites)			
30	Yes	Yes	204.0484	1	0.004	2.4	$C_8H_{10}FNO_2S$	AEBSF (protease inhibitor)		CID 1701	
31	Yes	No	489.4360	1	<0.1	<10	—				
32	Yes	Yes	152.1072	1	0.058	−1.1	$C_9H_{13}NO$	N-Methyltyramine / N-Methylphenylethanolamine	Amino alcohol / Amino alcohol	HMDB03633 HMDB01387	hsa00350

70

Table 1. Continued

No.	ABN	MeOH	m/z	z	Sigma™	Error (ppm)	Formula	Tentative identification	Classification	Database	KEGG pathway
33	Yes	Yes	218.0661	1	0.020	−0.6	$C_8H_{11}NO_6$	2-(Acetamidomethylidene)-3-(hydroxymethyl)butanedioic acid			
34	Yes	Yes	130.1592	1	0.007	−1.3	$C_8H_{19}N$	X			
35	Yes	No	228.1006	1	<0.1	<10	—				
36	Yes	No	230.0851	1	<0.1	<10	—				
37	Yes	Yes	244.0950	1	0.574	−9.0	$C_9H_{13}N_3O_5$	Cytidine	Nucleoside	HMDB00089	hsa0240
38	No	Yes	132.0771	1	0.007	−2.6	$C_4H_9N_3O_2$	Creatine β-Guanidinopropionic acid		HMDB00064 HMDB13222	hsa00260, hsa0330
39	No	Yes	264.0527	1	0.026	3.4	$C_9H_{13}NO_6S$	Epinephrine sulfate N-acetyl-S-(3-oxo-3-carboxy-n-propyl)cysteine		HMDB01876 HMDB02194	
40	Yes	Yes	303.6563	2	<0.1	<10	—	X			
41	Yes	No	188.2022	1	0.014	−1.8	$C_{11}H_{25}NO$	X			
42	Yes	No	174.1836	1	0.005	−5.6	$C_{10}H_{23}NO$	X			
43	No	Yes	218.1380	1	0.033	3.1	$C_{10}H_{19}NO_4$	Propionylcarnitine	Quaternary amine	HMDB00824	
44	Yes	Yes	192.1607	1	0.03	−1.4	$C_9H_{21}NO_3$	X			
45	Yes	No	271.1241	2	<0.1	<10	—				
46	Yes	Yes	548.1568	2	<0.1	<10	—				
47	Yes	Yes	200.2367	1	0.007	2.7	$C_{13}H_{29}N$	Octamylamine		CID 10406	
48	Yes	No	214.2522	1	0.008	3.5	$C_{14}H_{31}N$	X			
49	Yes	Yes	218.2083	1	0.133	2.2	$C_8H_{23}N_7$	X			
50	Yes	Yes	287.1039	1	0.010	6.4	$C_8H_{14}N_8O_2S$	X			
51	Yes	Yes	230.2475	1	0.02	1.7	$C_{14}H_{31}NO$	X			
52	Yes	Yes	118.0868	1	0.005	−4.4	$C_5H_{11}NO_2$	Valine (+3 metabolites)	Amino acid	HMDB00883[c]	hsa00640, hsa00770, hsa00280, hsa00290, hsa00970
53	Yes	Yes	106.0499	1	0.021	0.2	$C_3H_7NO_3$	Serine	Amino acid	HMDB00187[c]	hsa00600, hsa00920, hsa00460, hsa00270, hsa00260, hsa00970
54	No	Yes	301.1172	1	<0.1	<10	—	X			
55	Yes	Yes	387.0834	1	0.019	2.3	$C_{12}H_{10}N_{12}O_2S_2$				
56	Yes	Yes	349.1170	1	0.018	−7.8	$C_{15}H_{16}N_4O_6$	Riboflavin reduced	Pterin	HMDB01557	
57	Yes	Yes	132.1015	1	0.004	2.8	$C_6H_{13}NO_2$	Leucine or Isoleucine (+3 metabolites)	Amino acid	HMDB00687, HMDB00172[c]	hsa00280, hsa00290, hsa00970
58	No	Yes	299.1036	1	<0.1	<10	—				
59	Yes	No	335.1377	1	0.024	1.9	$C_{12}H_{22}N_4O_5S$	Tripeptide (V, N, C; M, Q, G; A, M, N)	Polypeptide	METLIN	
60	Yes	No	429.1509	1	<0.1	<10	—				
61	Yes	No	363.1681	1	0.023	4.5	$C_{14}H_{26}N_4O_5S_1$	Tripeptide (I/L,C,Q; N,V,M)	Polypeptide	METLIN	
62	Yes	No	258.2788	1	0.011	1.3	$C_{16}H_{35}NO$	X			
63	Yes	No	288.2892	1	0.037	1.9	$C_{17}H_{37}NO_2$	X			

Table 1. Continued

No.	ABN	MeOH	m/z	z	Sigma™	Error (ppm)	Formula	Tentative identification	Classification	Database	KEGG pathway
64	No	Yes	120.0659	1	0.737	−3.5	$C_4H_9NO_3$	Threonine (+2 metabolites)	Amino acid	HMDB00167[c]	hsa00860, hsa00290, hsa00260, hsa00970
65	Yes	Yes	291.1682	1	0.030	−6.4	$C_{11}H_{22}N_4O_5$	Tripeptide (G,S,K)	Polypeptide	METLIN	
66	No	Yes	147.0758	1	0.006	4.1	$C_5H_{10}N_2O_3$	Glutamine (+2 metabolites)	Amino acid	HMDB00641[c]	hsa00240, hsa04964, hsa00230, hsa00471, hsa00910, hsa00250, hsa00330, hsa02010, hsa00970
67	Yes	Yes	323.1961	1	0.018	1.3	$C_{17}H_{26}N_2O_4$	X			
68	Yes	No	305.1858	1	0.012	5.9	$C_{15}H_{29}O_4P$	Dolichol phosphate	Acyl phosphate	HMDB06353	hsa00510
69	Yes	No	351.2254	1	<0.1	<10	—	—			
70	No	Yes	116.0714	1	0.0029	−6.9	$C_5H_9NO_2$	Proline (+1 metabolite)	Amino acid	HMDB00162[c]	hsa00330, hsa02010, hsa00970
71	Yes	Yes	309.1806	1	0.009	0.9	$C_{16}H_{24}N_2O_4$	Bestatin (protease inhibitor)		CID 72172	
72	Yes	Yes	323.1964	1	0.110	0.6	$C_{17}H_{26}N_2O_4$	X			
73	No	Yes	365.0834	1	<0.1	<10	—	—			
74	Yes	Yes	327.1417	1	0.1885	−2.2	$C_{12}H_{18}N_6O_5$	Tripeptide (N,H,G)	Polypeptide	METLIN	
75	Yes	No	408.1959	1	<0.1	<10	—	—			
76	Yes	Yes	380.6276	2	<0.1	<10	—	—			
77	Yes	Yes	474.1998	1	<0.1	<10	—	—			
78	No	Yes	182.0819	1	0.035	−4.2	$C_9H_{11}NO_3$	Tyrosine	Amino acid	HMDB00158[c]	hsa00130, hsa00360, hsa00400, hsa00730, hsa00410, hsa00970
79	Yes	Yes	570.1186	1	<0.1	<10	—	—			
80	Yes	Yes	397.2848	1	<0.1	<10	—	—			
81	Yes	No	584.1268	1	<0.1	<10	—	—			
82	No	Yes	191.1025	1	0.005	0.6	$C_7H_{14}N_2O_4$	Diaminopimelic acid	Amino acid	HMDB01370	hsa00300
83	Yes	Yes	474.1990	1	<0.1	<10	—	—			
84	Yes	Yes	413.2810	1	<0.1	<10	—	—			
85	No	Yes	134.0451	1	0.0055	−2.5	$C_4H_7NO_4$	Aspartic acid (+1 metabolite)	Amino acid	HMDB00191[c]	hsa00760, hsa04080, hsa00340, hsa00410, hsa00460, hsa00770, hsa00910, hsa00250, hsa00270, hsa00300, hsa00260, hsa00330, hsa02010, hsa00970
86	Yes	Yes	429.3182	1	0.004	0.3	$C_{20}H_{40}N_6O_4$	X			
87	Yes	Yes	445.3136	1	<0.1	<10	—	—			
88	Yes	Yes	492.2153	1	0.028	−4.9	$C_{27}H_{29}N_3O_6$	Tripeptide (Y,Y,F)	Polypeptide	METLIN	
89	No	Yes	510.1815	1	0.159	2.1	$C_{21}H_{36}NO_9S_2$	X			
90	Yes	Yes	443.2985	1	<0.1	<10	—				
I.S.	No	Yes	182.0478	1	0.005	1.9	$C_5H_{11}NO_4S$	Methionine sulfone			
91	No	Yes	394.1854	1	0.106	−6.3	$C_{16}H_{31}N_3O_4S_2$	Tripeptide (I/L,M,M)	Polypeptide	METLIN	
92	No	Yes	307.0835	2	0.016	−0.8	$C_{20}H_{32}N_6O_{12}S_2$	Glutathione disulfide	Polypeptide	HMDB03337	hsa0480
93	No	Yes	400.1326	1	0.003	−3.8	$C_{11}H_{11}N_5O_{11}$	X			
94	No	Yes	396.1639	1	<0.1	<10	—	—			
95	No	Yes	290.1019	2	<0.1	<10	—	—			

Table 1. Continued

No.	ABN	MeOH	m/z	z	Sigma™	Error (ppm)	Formula	Tentative identification	Classification	Database	KEGG pathway
96	Yes	Yes	298.1022	2	<0.1	<10	–				
97	Yes	Yes	389.6241	2	<0.1	<10	–				
98	Yes	Yes	381.6264	2	<0.1	<10	–				
99	No	Yes	326.6199	2	<0.1	<10	–				
100	Yes	Yes	358.2086	1	0.024	–0.3	$C_{15}H_{27}N_5O_5$	E-64 (protease inhibitor)		CID 123985	
101	No	Yes	258.2770	1		+10					
102	No	Yes	326.6191	1		+10					
103	No	Yes	134.0455	1	0.017	–5.6	$C_4H_7NO_4$	Iminodiacetate			
104	No	Yes	217.0822	1	<0.1	<10	–				
105	No	Yes	186.0584	1	0.005	–0.60	$C_8H_{11}NO_2S$	X			
106	Yes	Yes	369.0941	1	<0.1	<10	–				
107	Yes	Yes	552.1293	1	<0.1	<10	–				
108	No	Yes	348.0731	1	0.012	–7.8	$C_{10}H_{14}N_5O_7P$	2′-Deoxyguanosine 5′-monophosphate Adenosine 2′-phosphate Adenosine 5′-phosphate	Nucleotide	HMDB01044 HMDB11617 HMDB00045	hsa01100[d]
109	No	Yes	184.0743	1	0.003	–9.6	$C_5H_{14}NO_4P$	Phosphocholine		METLIN	
110	Yes	Yes	478.1304	1	<0.1	<10	–				
111	Yes	No	516.0779	1	<0.1	<10	–				
112	Yes	No	529.3531	1	<0.1	<10	–				
113	No	Yes	293.0966	1	<0.1	<10	–				
114	No	Yes	504.1329	1	<0.1	<10	–				
115	No	Yes	212.0426	1	0.034	2.1	$C_4H_{10}N_3O_5P$	Phosphocreatine	Amino acid phosphate	HMDB01511	hsa01100[d]
I.S.	Yes	Yes	303.0607	1	0.006	2.7	$C_8H_{18}N_2O_6S_2$	Piperazinediethanesulfonic acid			

a) More than ten molecular formulas were assigned to the obtained *m/z* value.
b) No endogenous metabolite can be associated to the assigned molecular formula.
c) Tentative identification was confirmed with commercial standards.
d) hsa01100 code includes more than 20 metabolic pathways.

some trend (although not definitive) to low polarity amino acids when ABN cartridges are used. The rest of amino acids (arginine, lysine, GABA, threonine, glutamine, proline, tyrosine and aspartic acid) were only detected when methanol was used for purification. Other identified endogenous metabolites with high number of polar groups in their structure, such as gluthathione disulfide (peak 92), iminodiacetate (peak 103), AMP nucleotide (peak 108) and phosphocreatine (peak 115), could only be detected in the methanolic extract. Finally, most metabolites with phosphorus, sulfur or at least four oxygen atoms presented the highest migration times in good agreement with their expected electrophoretic mobilities at the separation pH. Moreover, a good number of tripeptides (peaks 8, 12, 14, 60, 61, 65, 74, 88 and 91) were proposed. Due to the lack of information from MS/MS spectra in this work, only amino acidic composition is indicated for each peptide.

The presence of three exogenous compounds was also observed, namely 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), bestatin and *N*-[*N*-(L-3-trans-carboxyoxirane-2-carbonyl)-L-leucyl]-agmatine (E-64) coming from the protease inhibitor cocktail used during cytosolic content isolation protocol. This finding is an additional demonstration on the consistency of these results.

In resume, from the 115 detected compounds, 44 metabolites were tentatively identified and classified in several groups. Most abundant compounds were amino acid and polypeptides, although other compounds such as lipoic acid derivatives, steroid derivatives, carbohydrates derivatives, amino alcohol, amino acid phosphates, nucleosides, quaternary amines, pterins, acyl phosphates and nucleotides were also found. Identified metabolites were automatically mapped to KEGG pathways using MassTrix. Among all tentatively identified compounds, 20 of them were related to the following metabolic pathways: (i) glycan biosynthesis and metabolism, (ii) organismal systems: excretory system, (iii) neuroactive ligand–receptor interaction, (iv) carbohydrate metabolism, (v) lipid metabolism, (vi) nucleotide metabolism, (vii) energy metabolism, (viii) metabolism of cofactors and vitamins, (ix) environmental information processing, (x) genetic information processing and (xi) amino acid metabolism. A particular case was creatine (peak 38), detected only in the methanolic extract, which was associated to more than 100 metabolic pathways. In Fig. 3, a bar plot representing the number of metabolites associated to the above-mentioned metabolic pathways, is presented. Information from both ABN and methanolic extracts is presented in order to show a general overview of the metabolic information obtained depending on the selected metabolite purification approach. As can be seen in Fig. 3, metabolites from both ABN and methanolic extracts were associated to nine metabolic pathways (numbered from 3 to 11), although the number of metabolites was different in practically all cases. For instance, in the metabolic pathway number 11 of Fig. 3, 16 metabolites were found in methanolic extract and only five in ABN extract. As a general trend
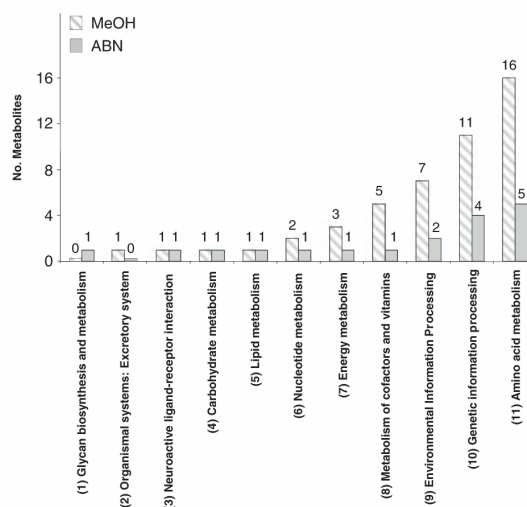
Figure 3. Number of compounds identified in some metabolic pathways in metabolite extracts obtained from SPE (ABN cartridge) or methanol deproteinization.

in this study, the use of methanol deproteinization brought about a higher number of metabolites associated to known metabolic pathways and, therefore, wider metabolomic information could be obtained.

4   Concluding remarks

Sample preparation is frequently underestimated in most metabolomic works. We have unequivocally demonstrated in this work that the composition and the quantity of metabolites detected depend to a large extent on the sample preparation step. Metabolite purification through methanol extraction showed good potential for metabolome characterization of human HT29 colon cancer cells, while the SPE results showed good extraction efficiency with different selectivity compared to protein methanol precipitation and ultrafiltration. However, SPE is usually considered more attractive for the on-line coupling extraction with CE-MS since sample manipulation can simplify overall analytical procedure, allowing automatization of the sample treatment, separation and detection in metabolomic studies. Selection of an appropriate sample treatment for a certain metabolomic study is, therefore, crucial. These results show the important influence from the metabolite purification strategy since it can bias and in some cases mislead the conclusions achieved by metabolomics. Hence, the question remains: is metabolomics approachable? Clearly, to achieve a whole metabolomics study of a biological system is still challenging at least based on the current available methodologies. A possible solution could be the use of multiple sample preparation procedures to cover a broad (and, therefore, more informative and representative) range of metabolites and concentrations.

5 References

[1] Shulaev, V., *Brief. Bioinform.* 2006, *7*, 128–139.

[2] Dettmer, K., Aronov, P. A., Hammock, B. D., *Mass Spectrom. Rev.* 2007, *26*, 51–78.

[3] Xiayan, L., Legido-Quigley, C., *Electrophoresis* 2008, *29*, 3724–3736.

[4] Garcia, D. E., Baidoo, E. E., Benke, P. I., Pingitore, F., Tang, Y. J., Villa, S., Keasling, J. D., *Curr. Opin. Microbiol.* 2008, *11*, 233–239.

[5] Issaq, H. J., Van, Q. N., Waybright, T. J., Muschik, G. M., Veenstra, T. D., *J. Sep. Sci.* 2009, *32*, 2183–2199.

[6] Monton, M. R. N., Soga, T., *J. Chromatogr. A* 2007, *1168*, 237–246.

[7] Harada, K., Fukusaki, E., *Plant Biotechnol.* 2009, *26*, 47–52.

[8] Ramautar, R., Somsen, G. W., de Jong, G. J., *Electrophoresis* 2009, *30*, 276–291.

[9] Ramautar, R., Mayboroda, O. A., Somsen, G. W., de Jong, G. J., *Electrophoresis* 2011, *32*, 52–65.

[10] Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., Forsythe, I., *Nucleic Acids Res.* 2009, *37*, D603–D610.

[11] Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., Proceedings of the 9th International Congress of Therapeutic Drug Monitoring and Clinical Toxicology, Louisville, Kentucky, April 23–28, 2005.

[12] Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M., *Nucleic Acids Res.* 2010, *38*, D355–D360.

[13] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M., *Nucleic Acids Res.* 2006, *34*, D354–D357.

[14] Kanehisa, M., Goto, S., *Nucleic Acids Res.* 2000, *28*, 27–30.

[15] Suhre, K., Schmitt-Kopplin, P., *Nucleic Acids Res.* 2008, *36*, W481–W484.

[16] Timerbaev, A. R., Hirokawa, T., *Electrophoresis* 2006, *27*, 323–340.

SUPPORTING INFORMATION
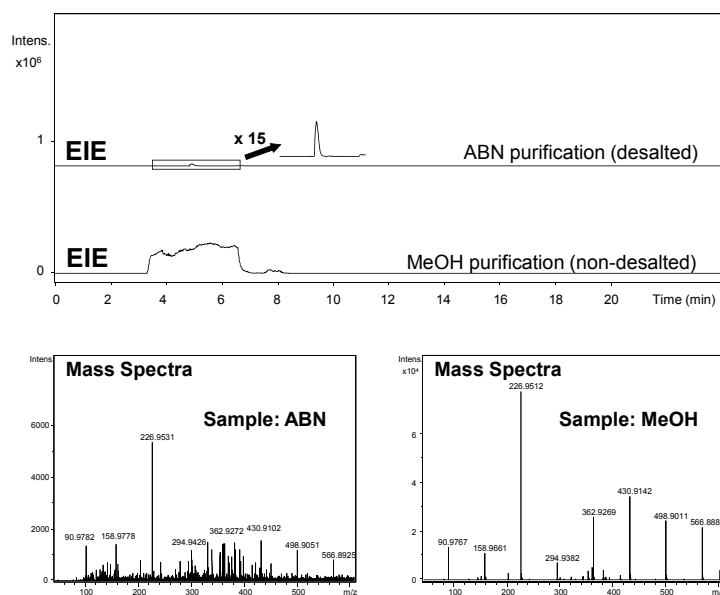
Figure S1. CE-MS extracted ion electropherograms of the sodium formate clusters obtained from ABN extract (upper electropherogram) and methanolic extract (lower electropherogram). Characteristic mass spectra are shown in the lower panel. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section.
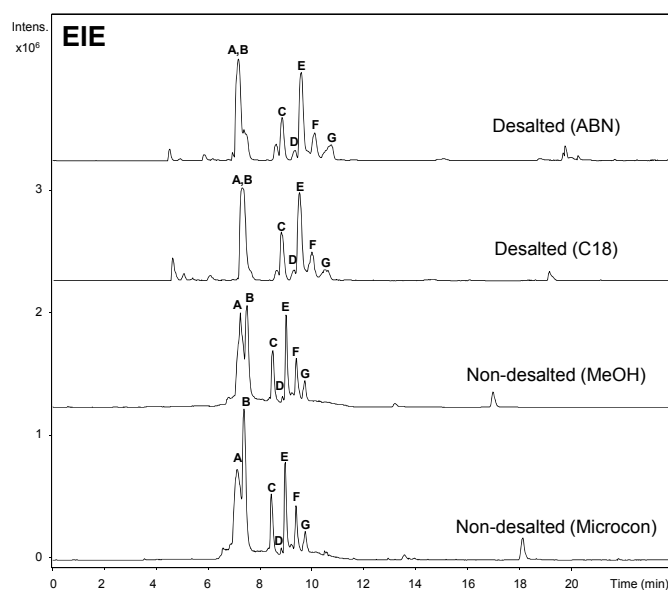
Figure S2. CE-MS extracted ion electropherograms of the 43 m/z ions common to all extracts. Seven peaks (from A to G) are marked in the electropherograms for a better comparison of the purification strategies. Peak A: 122.08 m/z, peak B: 204.05 m/z, peak C: 387.08 m/z, peak D: 291.17 m/z, peak E: 309.18 m/z, F: 413.29 m/z, and G: 182.05 m/z. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section. ase peak electropherograms of the cytosolic fraction from colon cancer cell culture after four different metabolite extraction procedures. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section. Esquema de la interfase electrospray con flujo adicional.
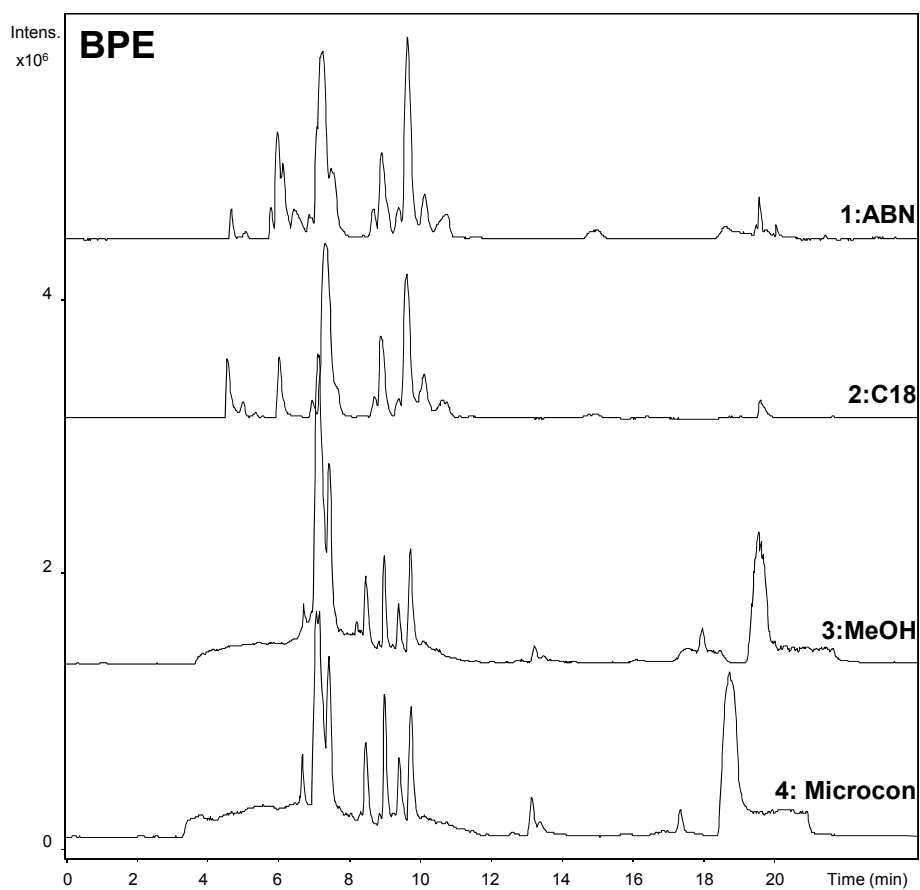
Figure S3. Base peak electropherograms of the cytosolic fraction from colon cancer cell culture after four different metabolite extraction procedures. CE-MS conditions are described in "CE-ESI-TOF MS conditions" section. Esquema de la interfase electrospray con flujo adicional.

Table S1. Peak areas, migration times and RSD values for ten different metabolites detected by CE-MS after ABN extraction in three different days. CE-MS conditions are described in section 2.5.

DAY 1	R1		R2		R3		R4		R5		t_mig			Area		
Metabolite No.	t mig	A	t mig	A	t mig	A	t mig	A	t mig	A	Med	SD	RSD	Med	SD	RSD
15	6.55	8265	6.62	7654	6.75	8038	6.85	7652	6.97	8017	6.75	0.17	2.5	7925	267	3.4
22	7.05	78569	7.15	72086	7.29	74241	7.37	75540	7.50	82332	7.27	0.18	2.4	76554	3993	5.2
32	7.57	725663	7.69	702632	7.86	767788	7.95	715422	8.09	763530	7.83	0.21	2.6	735007	29186	4.0
40	8.16	14532	8.22	14628	8.45	15550	8.55	15201	8.71	14123	8.42	0.23	2.7	14807	566	3.8
46	8.70	4205	8.77	3955	8.93	4475	9.01	4503	9.19	4355	8.92	0.20	2.2	4299	225	5.2
56	9.40	110523	9.46	116568	9.65	112865	9.76	102568	9.94	109936	9.64	0.22	2.3	110492	5139	4.7
72	10.35	2685020	10.41	2638364	10.58	2890925	10.71	2769980	10.88	2432013	10.59	0.22	2.1	2683260	170139	6.3
88	11.15	130982	11.23	127066	11.39	141619	11.52	146230	11.68	125969	11.39	0.21	1.9	134373	9064	6.7
91	11.51	2664120	11.70	2874057	11.79	2859792	11.89	2669031	12.16	2508991	11.81	0.24	2.0	2715198	152811	5.6
106	16.11	335642	17.53	308625	17.40	381204	16.90	370219	17.91	359821	17.17	0.70	4.0	351102	29116	8.3

DAY 2	R1		R2		R3		R4		R5		t_mig			Area		
Metabolite No.	t mig	A	t mig	A	t mig	A	t mig	A	t mig	A	Med	SD	RSD	Med	SD	RSD
15	6.12	8803	6.22	9288	6.32	8166	6.44	8456	6.69	9113	6.36	0.22	3.5	8765	461	5.3
22	6.65	65684	6.63	68612	6.79	59917	6.93	62464	7.21	64121	6.84	0.24	3.5	64160	3280	5.1
32	7.24	700996	7.17	741936	7.35	643436	7.55	675458	7.85	693140	7.43	0.27	3.7	690993	36065	5.2
40	7.79	17698	7.62	16398	7.86	14854	8.08	16584	8.40	16918	7.95	0.30	3.8	16490	1041	6.3
46	8.28	4809	8.00	5034	8.29	4350	8.54	4583	8.87	4664	8.40	0.33	3.9	4688	255	5.4
56	8.90	95684	8.52	98724	8.85	85245	9.14	89878	9.49	91184	8.98	0.36	4.0	92143	5232	5.7
72	9.75	2515102	9.26	2587314	9.65	2220188	9.97	2355491	10.33	2384463	9.79	0.40	4.0	2412512	143272	5.9

Table S1. (cont.)

	R1		R2		R3		R4		R5		t_{mig}			Area		
	t mig	A	t mig	A	t mig	A	t mig	A	t mig	A	Med	SD	RSD	Med	SD	RSD
88	10.42	158847	9.82	167502	10.28	143237	10.63	152494	11.00	153825	10.43	0.44	4.2	155181	8902	5.7
91	10.86	2375011	10.19	2488590	10.70	2112414	11.05	2265612	11.43	2278126	10.85	0.46	4.2	2303951	139610	6.1
106	14.82	434226	13.92	452856	14.65	379092	15.10	412280	15.58	410724	14.81	0.61	4.1	417836	27738	6.6

DAY 3	R1		R2		R3		R4		R5		t_{mig}			Area		
Metabolite No.	t mig	A	t mig	A	t mig	A	t mig	A	t mig	A	Med	SD	RSD	Med	SD	RSD
15	5.85	9566	5.92	10005	6.02	9259	6.08	10082	6.25	9577	6.02	0.15	2.6	9698	342	3.5
22	6.33	63332	6.35	59379	6.43	56421	6.50	63777	6.67	59206	6.46	0.14	2.1	60423	3094	5.1
32	6.82	829796	6.87	880844	6.98	781408	7.00	846435	7.24	808650	6.98	0.16	2.3	829427	37652	4.5
40	7.24	16104	7.34	17867	7.42	17621	7.44	18864	7.69	17972	7.43	0.17	2.3	17686	1001	5.7
46	7.58	4991	7.74	4961	7.77	4492	7.80	5002	8.07	4853	7.79	0.18	2.3	4860	214	4.4
56	8.09	86017	8.31	85230	8.32	94152	8.32	82006	8.62	86626	8.33	0.19	2.3	86806	4476	5.2
72	8.72	2179522	9.07	2305314	8.99	2083717	8.96	2208366	9.35	2097898	9.02	0.23	2.5	2174963	89990	4.1
88	9.27	155015	9.70	158715	9.57	175790	9.52	177548	9.94	166363	9.60	0.25	2.6	166686	10009	6.0
91	9.59	2165933	10.05	2021198	9.89	2326637	9.84	2040365	10.30	2149446	9.93	0.26	2.6	2140716	122100	5.7
106	13.67	402900	14.48	437380	14.17	481275	13.92	454391	14.71	453070	14.19	0.42	2.9	445803	28707	6.4

Table S2. Overall reproducibility (sample preparation + CE-MS analysis) given as peak areas and RSD values for metabolites 56, 72 and 88 determined in three ABN and MeOH extracts obtained under the same conditions and analyzed in triplicate by CE-MS (total n=9).

No.	m/z	CE-MS Rep.	Peak area (Treatment 1)	Peak area (Treatment 2)	Peak area (Treatment 3)	Mean	SD	% RSD
				ABN				
56	349.117	R1	105065	101583	108973	104504	6462	6
		R2	107374	92001	113464			
		R3	104004	98574	109500			
72	323.1964	R1	2296328	2359709	2465852	2395546	154976	6
		R2	2424419	2135371	2608870			
		R3	2389613	2269479	2610277			
88	492.2153	R1	130842	154736	162406	156981	18451	12
		R2	141178	154733	175277			
		R3	139635	164498	189523			
				MeOH				
56	349.117	R1	23018	26045	24145	24658	1852	8
		R2	21283	27317	24878			
		R3	23553	25745	25939			
72	323.1964	R1	10958	10716	12283	10949	1129	10
		R2	9939	10175	11909			
		R3	9862	9879	12819			
88	492.2153	R1	130118	102388	95395	110122	14524	13
		R2	118505	96908	102439			
		R3	135152	101220	108975			

3.3. CAPÍTULO 2 (Chapter 2)

CE/LC-MS Multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation

Clara Ibáñez[1]
Carolina Simó[1]
Virginia García-Cañas[1]
Ángeles Gómez-Martínez[2]
José A. Ferragut[2]
Alejandro Cifuentes[1]

[1]Laboratory of Foodomics, CIAL (CSIC), Madrid, Spain
[2]Institute of Molecular and Cellular Biology, Miguel Hernández University, Avda. Universidad s/n, Elche, Alicante, Spain

Research Article

CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation

In this study, an analytical multiplatform is presented to carry out a broad metabolomic study on the anti-proliferative effect of dietary polyphenols on human colon cancer cells. CE, RP/UPLC, and HILIC/UPLC all coupled to TOF MS were combined to achieve a global metabolomic examination of the effect of dietary polyphenols on HT29 colon cancer cells. By the use of a nontargeted metabolomic approach, metabolites showing significant different expression after the polyphenols treatment were identified in colon cancer cells. It was demonstrated that this multianalytical platform provided extensive metabolic information and coverage due to its complementary nature. Differences observed in metabolic profiles from CE-TOF MS, RP/UPLC-TOF MS, and HILIC/UPLC-TOF MS can be mainly assigned to their different separation mechanisms without discarding the influence of the different tools used for data processing. Changes in glutathione metabolism with an enhanced reduced glutathione/oxidized glutathione (GSH/GSSG) ratio were detected in polyphenols-treated cells. Moreover, significant alterations in polyamines content with important implications in cancer proliferation were observed after the treatment with polyphenols. These results from metabolomics can explain the chemopreventive effect of the tested dietary polyphenols on colon cancer and may be of importance for future prevention and/or treatment of this disease.

Keywords:

CE-MS / Colon cancer / Metabolomics / Polyphenols / UPLC-MS

DOI 10.1002/elps.201200143

1 Introduction

It is not clear how many metabolites constitute the human metabolome, but it has been estimated that there are approximately 3000 metabolites [1]. The task of simultaneously profiling all metabolites from a biological system is challenging due to their different physicochemical proper-

Correspondence: Dr. Carolina Simó, Laboratory of Foodomics, CIAL (CSIC), Nicolás Cabrera 9, 28049 Madrid, Spain.
E-mail: c.simo@csic.es
Fax: +34-910-017905

Abbreviations: GSH, reduced glutathione; **GSSG**, oxidized glutathione; **IPA**, Ingenuity pathways analysis; **KEGG**, Kyoto encyclopedia of genes and genomes; **METLIN**, Metabolite mass spectral database; **MH**, MassHunter; **MPP**, Mass profiler professional; **SSAT**, spermidine/spermine N-acetyltransferase; **UPLC**, ultra-high-performance liquid chromatography

ties, varying in terms of molecular mass, polarity, effective charge, and wide dynamic range of concentrations. Besides, the choice of an adequate sample preparation step has already been shown to be crucial in any metabolomic study [2]. Actually, it is now assumed that the coverage of the human metabolome is impossible to achieve with a single analytical methodology. Typical metabolomic studies are based on MS or NMR. Methods based on NMR have been proved to be robust and reproducible in metabolomics, although the low NMR sensitivity only allows obtaining information on the most abundant metabolites. On the other hand, metabolites can be identified by NMR chemical shift measurements [3, 4]. MS presents higher sensitivity compared to NMR and it is being increasingly used as an alternative promising approach for metabolomics. However, metabolites do not ionize to an equal extent during MS analysis, biasing in this way the information produced. MS is usually coupled to a chromatographic/electrophoretic separation technique such as LC, GC, or CE to enhance resolution and sensitivity. CE is a fast and high-resolution separation technique. Although

CE has been demonstrated to be an interesting and useful tool in metabolomics [5–10], compared to GC or LC, it has not been extensively used to carry out metabolomic studies. Compared to GC-MS, both LC-MS and CE-MS can make possible the analysis of higher polarity compounds and lower volatility in a wide mass range without derivatization. In this regard, HILIC is becoming a complementary analytical mode to the more common RP/LC also in metabolomics, due to the ability of HILIC to separate more hydrophilic metabolites. Retention mechanisms in HILIC are a combination of hydrophilic interaction, ion-exchange, and reversed-phase retention that result in enhanced retention of polar analytes. On the other hand, the use of high and ultra-high resolution mass analyzers (TOF, FTMS, Orbitrap®, etc.) is essential to obtain accurate mass measurements for the determination of elemental compositions of metabolites and to carry out tentative identification based on metabolites databases. Moreover, metabolomics generates large amounts of data, making data processing a critical step for any application in this field. Preprocessing and statistical analysis of these data, involve significant challenges and require specialized bioinformatics tools. The existing and emerging computational algorithms for data processing are a key component. In this regard, when the objective is to detect as many different metabolites as possible, raw data processing is a very important step in data analysis. Main data processing workflow steps include noise reduction, spectrum deconvolution, electropherogram/chromatogram alignment and peak integration. However, it is interesting to remark that currently there is no standard data processing workflow for CE/LC-MS-based metabolomics.

Metabolomics has been predominantly applied to discover biomarkers related to prognosis, diagnosis, and therapeutic monitoring of diseases including cancer, since significant metabolic changes take place as cells are transformed from normal to malignant [11]. In addition to conventional therapeutic drugs, numerous natural dietary constituents are now under scrutiny due to their promising anti-cancer properties [12–14]. Among dietary constituents, polyphenols have been claimed to show promising anticancer activities [15–18]. Although many of the health benefits assigned to many dietary constituents are still under controversy, it is clear that more sounded scientific evidences will help to elucidate their claimed beneficial effects as also proposed by the new foodomics discipline [19]. As an additional consequence, better scientific evidences will allow an easier approval by food authorities of these compounds (e.g. as new functional ingredients). In this sense, the advent of new postgenomic technologies will be essential to understand and explain how the bioactive compounds from diet interact at molecular and cellular level. Among these new technologies, metabolomics will be a crucial tool to understand the molecular mechanisms of bioactive compounds in health/disease studies.

In the present work, a multiplatform based on CE-TOF MS, RP/UPLC-TOF MS, and HILIC/UPLC-TOF MS is pre-sented as an effective strategy to characterize the metabolome changes in colon cancer cells whose proliferation was extensively reduced by a rosemary extract rich in polyphenols. This study is based on a hypothesis-free approach and shows that the combination of advanced analytical techniques with bioinformatics and statistical tools provide potential metabolic markers that can serve as indicators of disease progression or response to diet intervention.

2 Materials and methods

2.1 Chemicals

All chemicals were of analytical reagent grade and used as received. Reagents and solvents employed in the preparation of LC mobile phases, CE electrolytes, sheath liquid, and standard solutions were of MS grade: formic acid and 2-propanol were from Riedel-de Haën (Seelze, Germany), ACN and water were from Labscan (Gliwice, Poland). Metabolite standards were from Sigma-Aldrich (St. Louis, MO, USA).

2.2 Polyphenol extract

Extracts rich in dietary polyphenols were obtained from rosemary (*Rosmarinus officinalis*) using supercritical fluid extraction (SFE) under the following conditions of pressure and solvent composition: supercritical CO_2 plus 7% ethanol (v/v) at 150 bar and 40°C for 300 min were used as previously reported by Herrero et al. [20]. Chemical characterization of this extract provided information on the major phenolic constituents, namely, cirsimaritin, genkwanin, carnosol, carnosic acid, and methyl carnosate. This extract was reported to contain high concentrations of carnosol (226.39 μg/mg extract) and carnosic acid (151.55 μg/mg extract). The extract has also been reported to show antioxidant activity (EC_{50} value of 8.1 μg/mL).

2.3 Cell culture and growth conditions

HT29 cells (from Institut Municipal d´Investigació Médica, Barcelona, Spain) were grown in DMEM supplemented with 5% (v/v) heat-inactivated fetal calf serum, 2 mM of L-glutamine, 50 U/mL of penicillin G, and 50 μg/mL of streptomycin, at 37°C in humidified atmosphere and 5% CO_2 (v/v). Cells were plated at a density of 10 000 cells/cm^2 on 60 mm diameter culture plates and exposed for 72 h to a dissolution containing a total concentration of 10 μM in the mentioned rosemary phenols (i.e. treated cells). The respective control cells were grown under identical conditions but without the rosemary phenols (i.e. control cells).

2.4 Cell cycle study

The nuclear DNA content was quantitatively measured at high speed by flow cytometry. Thus, control and polyphenol-treated HT29 cells were trypsinized, washed with PBS, and fixed with 75% (v/v) cold ethanol at −20°C for at least 1 h. Cells were then incubated with 0.5 Triton X-100 and 25 μg/mL RNase A in PBS, stained with 25 ng/mL propidium iodide and analyzed using an Epics XL flow cytometer equipped with a 0.75W argon laser set at 488 nm (Beckman Coulter Co., Miami, FL, USA). Accumulation of HT29 cells in G2/M phase and a corresponding reduction in the proportion of cells in G1 phase was observed in polyphenol-treated cells. A pro-apoptotic effect supported by a slight accumulation of subG1 populations of HT29 cells was also observed.

2.5 Metabolite extraction procedure

A PBS solution containing 138 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM sodium hydrogen phosphate at pH 7.4, was purchased from Sigma-Aldrich. Homogenization buffer (10 mM Tris-HCl, 5 mM EDTA, 120 mM NaCl, at pH 7.4) and protease inhibitor cocktail (4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), pepstatin A, E-64, bestatin, leupeptin, and aprotinin, was also from Sigma-Aldrich. Equal volumes of human HT29 colon cancer cells were washed with PBS solution and centrifuged. The pellet was suspended with homogenization buffer and protease inhibitor cocktail. Cells were disrupted with a Polytron homogenizer and centrifuged (14 min at $14\ 000 \times g$ and 4°C). Pellet (mainly, nuclear and mitochondrial fraction) was discarded and supernatant was centrifuged for 1 h at $100\ 000 \times g$ and 4°C. The resulting supernatant (assigned as cytosolic fraction) was frozen at −80°C. Further metabolite purification was needed considering the relatively high amount of proteins present in the cytosolic fraction. Protein precipitation with methanol as well as ultrafiltration protocols, were found to be satisfactory and reproducible for metabolite purification from cytosolic content [2]. First, total protein concentration of cytosolic fraction was determined using a modified Bradford assay from Bio-Rad (Hercules, CA, USA). For metabolite purification 400 μL of cytosolic fraction were ultrafiltrated with an Amicon Ultra 3 kDa centrifugal device (from Millipore, Bedford, MA, USA) at $14\ 000 \times g$ and 20°C for 40 min.

2.6 CE-MS conditions for metabolite analysis

CE analyses were carried out in a P/ACE 5500 CE apparatus from Beckman Instruments (Fullerton, CA, USA). The instrument was controlled by a PC running the System Gold software from Beckman. CE was coupled to MS through an orthogonal ESI interface model G1607A from Agilent Technologies (Palo Alto, CA, USA). Electrical contact at the ESI needle tip was established via a sheath liquid delivered by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills,

IL, USA). A TOF MS instrument (model micrOTOF) from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a PC running the micrOTOF control software from Bruker Daltonics.

The electrophoretic separation was carried out using an uncoated fused-silica capillary (50 μm id, 363 μm od and 80 cm total length) from Composite Metal Services (Worcester, England). Before first use, the separation capillary was conditioned by rinsing with 1 M NaOH for 10 min, followed by 20 min with water both using pressurized N_2 at 20 psi (1380 mbar). After each run, the capillary was conditioned with water during 2 min, followed by BGE during 4 min. Injections were made at the anodic end using N_2 pressure at 0.5 psi (34.5 mbar) for 80 s. Samples were analyzed by quintuplicate. The electrophoretic separation was achieved applying +25 kV (30–32 μA current) at a constant temperature of 25°C in a BGE composed of 1 M formic acid. Electrical contact at the ESI needle tip was established via a sheath liquid based on isopropanol water (50:50, v/v) and delivered at a flow rate of 0.24 mL/min. The mass spectrometer operated in the positive ion mode. The nebulizer and drying gas conditions were 0.4 bar N_2 and 4 L/min N_2, respectively, and maintaining the ESI chamber temperature at 250°C. Spectra was acquired in the 50–600 m/z range every 90 ms. Data was processed using the DataAnalysis 4.0 software from Bruker Daltonics. External and internal calibration of the TOF MS instrument was performed by introducing a 10 mM sodium formate solution through the separation capillary. The ions used for the calibration of the TOF MS instrument were next: 90.9766, 158.9641, 226.9515, 294.9389, 362.9263, 430.9138, 498.9012, and 566.8886 m/z.

2.7 UPLC-MS conditions for metabolite analysis

The UPLC system Agilent 1290 (Agilent Technologies, Santa Clara, CA, USA) was connected to a Q/TOF MS Agilent 6540 (Agilent Technologies) equipped with an orthogonal ESI interface (Agilent Jet Stream, AJS) with Jet Stream thermal gradient focusing technology and operating in positive ion mode. The instrument was controlled by a PC running the Mass Hunter Workstation software from Agilent.

RP chromatographic separation was performed on an Agilent ZORBAX C18, Rapid Resolution HT (2.1 × 50 mm, 1.8 μm) column maintained at 40°C. Elution was performed using phase A (water with 0.1% formic acid, v/v) and phase B (acetonitrile with 0.1% formic acid, v/v) and the following gradient program: 2–20% B in 6 min, 20–100% B in 4 min. After analysis, 100% B was kept constant for 2 min more and then the column was re-equilibrated for 5 min using the initial solvent composition.

HILIC was performed on a ZORBAX HILIC Plus HT (2.1 × 50 mm, 1.8 μm) column maintained at 30°C. Elution was performed using phase A (water with 10 mM ammonium formate at pH 5) and phase B (acetonitrile) and the following gradient program: 90–80% B in 7 min, 80–50% B in 2 min, and 50–0% B in 1 min. Then, 0% B was kept constant for

2 min and finally 90% B was set during 5 min for column re-equilibration.

MS operation parameters were the following: capillary voltage, −4000 V; nebulizer pressure, 40 psi; drying gas flow rate, 10 L/min; gas temperature, 300°C; skimmer voltage, 45 V; fragmentor voltage was 125 V in positive mode. TOF MS accurate mass spectra were recorded across the range of 50–1000 *m/z* every 200 ms. Internal mass calibration of the instrument was carried out using an AJS ESI source with an automated calibrant delivery system. External calibration of the TOF MS was carried out using a commercial mixture from Agilent with next *m/z* values: 118.086255, 322.048121, 622.028960, 922.009798, 1221.990637, and 1521.971475. The reference compound solution for internal mass calibration of the Q/TOF mass spectrometer containing 5 µM of purine ([C5H5N4]$^+$ at 121.050873 *m/z*) and 2.5 µM HP-0921, hexakis (1H,1H,3H-tetrafluoropropoxy) phosphazine ([C18H19O6N3P3F24]$^+$ at 922.009798 *m/z*) in ACN/water (95:5, v/v) was also from Agilent.

2.8 Quality control

Quality control (QC) samples were used to monitor the performance of the method and increase the reliability of the data. QC was prepared by combining equal aliquots from each sample (control cells and polyphenol-treated cells) and injected regularly throughout the run to monitor the stability of the CE/UPLC-MS method.

2.9 Data processing

In this work, two different data processing workflows were used to compare multiple datasets. Raw CE-MS data were processed using the XCMS open-source package [21], written in the platform-independent programming language R (http://www.r-project.com). Raw data signals obtained from CE-MS analysis were converted into centroid peak data and exported to the open source MS exchange format netCDF using Data Analysis 4.0 (Bruker). netCDF files from CE-MS were subsequently processed using XCMS. Peak detection method used in XCMS was the CentWave algorithm [22], which finds a region of interest based on the mass accuracy and expected chromatographic/electrophoretic peak width. Found peaks were reported in the form of "features", which are defined as a unique m/z value at a unique time point. XCMS performs nonlinear retention time correction, using the OBI-Warp algorithm (Ordered Bijective Interpolated Warping) [23]. The resulting output data table from XCMS containing peak information across samples was used in R for subsequent peak area normalization. There is not a general ideal way of normalization in metabolomics. In this work, average of intensities based on complete dataset was used as scaling factor for each sample. Finally, a multidimensional data matrix of time-aligned detected features with their corresponding migration time, *m/z* and peak area was obtained

for each sample. Redundant responses from the same ion (isotopes, adducts, fragment ions, etc.) were detected using the R-software package AStream [24].

On the other hand, raw data signals obtained from UPLC-MS were analyzed using MassHunter software 4.0 (MH) and Mass Profiler Professional 2.2 (MPP), both from Agilent. Unlike XCMS, the MH and MPP software offered a graphical user interface. MH program uses a proprietary molecular feature extraction (MFE) algorithm to find compounds in the entire set of mass spectra from each chromatogram. This algorithm checks also the presence of adducts (Na$^+$, K$^+$, etc.), isotopes ([M+H]$^+$, [M+H+1]$^+$, etc.) dimers from the same compound, neutral losses (H$_2$O, NH$_2$, etc.). After MFE a Compound Exchange File (.cef) is generated for each sample analysis. This cef file is then exported to MPP software for retention time correction, area normalization, statistics, and visualization of results.

Both CE-MS and UPLC-MS data processing were carried out using a standard HP computer Z400 (2.8 GHz, 4 GB memory) running Windows 7.

2.10 Statistical analysis

Comparative analysis was performed by calculating fold changes together with their respective Welch *t*-test to assess the significance of the changes. Each ion was tested for its statistical significance. A two-way parametric Student's *t*-test was applied to compare the significance of changes in measured ion responses (normalized areas) for metabolites in cytosol of the treated cells compared to the control cells. Only differences with a critical probability value *p* < 0.05 were considered statistically significant.

2.11 Metabolite identification and pathway analysis

After statistical analysis of CE-MS datasets, significantly different molecular features between control and polyphenol-treated cells were subjected to identification process. Metabolite tentative identification was performed by matching the obtained accurate *m/z* values and theoretical *m/z* values contained in different free available databases, namely: Human Metabolome Database [25], METLIN [26], KEGG compound [27], and PubChem [28] with a mass accuracy window of 10 ppm. The metabolite tentative identification from both UPLC-MS datasets was integrated in the MPP software through the annotation of compounds of METLIN. Comparison of both UPLC-MS and CE-MS experimental masses with other databases different to METLIN was carried out ion by ion.

When available, standards were used to confirm metabolite identification by means of sample coinjection. Generate-MolecularFormula Editor within Data Analysis 4.0 (Bruker) and Molecular Formula Generator algorithm within MassHunter (Agilent) software were used to support the agreement, in terms of mass error (ppm) and comparison to the theoretical isotopic pattern, between the molecular

formula generated by the software and the proposed compound from metabolite database search. When isomers exist for a given formula, metabolite identification was sorted giving preference to metabolites from central metabolic pathways in KEGG and number of databases containing each metabolite.

Finally, metabolomic datasets were analyzed using Ingenuity Pathways Analysis, IPA[R], software from Ingenuity Systems (Redwood City, CA, USA). A dataset containing KEGG metabolite identifiers and the corresponding fold change values was uploaded into the IPA application.

3 Results

3.1 CE-ESI-TOF MS metabolic analysis of human HT29 colon cancer cells

CE-MS is mainly focused on the analysis of ionic, weakly ionic, and/or highly polar metabolites, which comprises a high proportion of the already known metabolites. At the acidic separation conditions (pH 2) selected in this work, most compounds are positively charged. Control and treated cells seem to provide very similar and simple CE-MS profiles. However, a deeper data analysis carried out using XCMS software gave more than 2890 features detected along the electropherogram. These features from control and treated groups were further subjected to data processing, filtering and statistical analysis as described in Sections 2.8 and 2.9. A final list was obtained with 212 compounds that showed statistically significant variations in their abundances as response to the treatment of colon cancer cells with polyphenols. Identification of the metabolites differentially expressed ($p < 0.05$) was performed as described in Section 2.11, providing the tentative identification of 22 compounds as shown in Table 1. Extracted ion electropherograms from the above-mentioned identified compounds are presented in Fig. 1.

3.2 RP/UPLC-ESI-TOF MS metabolic analysis of human HT29 colon cancer cells

The reduced particle size used in UPLC (1.8 μm) compared to normal particle size in HPLC (3.5-5 μm), results in increased peak capacity, resolution, and lower analysis time. This improvement in UPLC resolution is very relevant when using MS detection to avoid ion suppression related to coelution of close metabolites. After data analysis using Mass Hunter and Mass Profiler Professional software, more than 2176 initial features that appeared in 100% of replicates belonging to the same class of sample (control or polyphenol-treated cells) were detected along the chromatograms. After signal filtering, a total number of 544 different compounds were subjected to statistical analysis. Among the 544 detected compounds, 210 were statistically different with a significance level lower than 0.05. Among all significantly differ-

ent compounds observed, 32 were tentatively identified (see Table 1).

3.3 HILIC/UPLC-ESI-TOF MS metabolic analysis of human HT29 colon cancer cells

Ammonium formate at pH 5 was used as ionic additive to control the mobile phase pH and ion strength in HILIC/UPLC-MS analysis. It was observed that this mobile phase improved analyte desolvation and ionization, enhancing MS sensitivity. Indeed, higher sample dilution compared to reverse phase RP/UPLC-MS analysis of Section 3.2 was needed to avoid signal saturation. The workflow for data processing was the same as the one used for RP/UPLC-MS experiments. Under these conditions, 1077 features were initially observed. After signal filtering, 214 were found to be differently expressed ($p < 0.05$). Among all significantly different compounds 12 were tentatively identified (see Table 1).

4 Discussion

4.1 From peaks to metabolites

The metabolomic approach proposed in this work does not aim to identify the entire set of metabolites from cell cytosol, but rather to compare patterns of metabolites that change after treating colon cancer cells with rosemary polyphenols. In general, compound identification is a key part after statistical analysis in nontargeted metabolomic studies. A major difficulty in the identification process is the MS signal redundancy and artifacts. The presence of adducts and fragments, produced by CID in the transfer region of the mass spectrometer, hinders the mass spectra interpretation. Based on the statistical analysis, peaks with the highest statistical significance (p value < 0.05) were subjected to the identification process. From the total list of compounds with statistically significant differences among samples, a limited number of compounds could be tentatively identified (Table 1). Still today, qualitative identification of unknown metabolites represents one of the most substantial challenges in metabolomics, since databases remain incomplete and the access to commercial standards is limited. Among the 65 tentatively identified metabolites founded to be significantly different ($p < 0.05$) after the polyphenol treatment of colon cancer cells, 51 were determined to be significantly upregulated and 14 downregulated (Table 1).

4.2 CE-MS and UPLC-MS comparison

Electrospray is the interface most commonly used for metabolomic studies by CE/UPLC-MS-based strategies. When working with CE-MS, an additional difficulty has to be overcome owing to the need to close the CE electrical circuit. The use of coaxial sheath flow ESI interface improved

Liquid Phase Separations   2333

Table 1. Names and main properties of tentatively identified metabolites from HT29 colon cancer cells that showed significant differences ($p < 0.05$) after the treatment with rosemary polyphenols

Migration/ retention time	Measured m/z	Error (ppm)	Ion	Formula	Tentative identification	Fold change	Regulation
9.19[a]	188.1753	− 2.3	(M+H)	$C_9H_{21}N_3O$	N-Acetylspermidine	6.3	Up
10.66[a]	90.0545	− 5.9	(M+H)	$C_3H_7NO_2$	Sarcosine	2.6	Up
10.73[a]	175.1181	− 5.2	(M+H)	$C_6H_{14}N_4O_2$	Arginine[d]	3.3	Up
10.81[a]	156.0761	− 4.2	(M+H)	$C_6H_9N_3O_2$	Histidine[d]	6.0	Up
10.83[a]	147.1124	− 2.4	(M+H)	$C_6H_{14}N_2O_2$	Lysine[d]	3.6	Up
10.87[a]	145.1080	− 2.5	(M+H)	$C_5H_{12}N_4O$	4-Guanidinobutanamide	2.1	Up
12.34[a]	162.1124	− 0.6	(M+H)	$C_7H_{15}NO_3$	Carnitine [d]	2.8	Up
13.26[a]	385.1289	− 0.1	(M+H)	$C_{14}H_{20}N_6O_5S$	S-adenosyl-L-homocysteine	3.1	Up
13.50[a]	132.0777	6.9	(M+H)	$C_4H_9N_3O_2$	Creatine [d]	3.2	Up
14.21[a]	175.1064	− 7.3	(M+H)	$C_7H_{14}N_2O_3$	N-Acetyl-L-Ornithine	3.6	Up
14.36[a]	90.0542	− 8.9	(M+H)	$C_3H_7NO_2$	Alanine [d]	3.1	Up
14.69[a]	268.1031	− 3.2	(M+H)	$C_{10}H_{13}N_5O_4$	Adenosine	2.9	Up
15.14[a]	218.1143	3.5	(M+H)	$C_8H_{15}N_3O_4$	N-Acetyl-L-Citrulline	3.7	Up
15.35[a]	219.1343	2.1	(M+H)	$C_9H_{18}N_2O_4$	N2-(D-1-Carboxyethyl)-L-Lysine	3.3	Up
15.75[a]	118.0868	4.2	(M+H)	$C_5H_{11}NO_2$	Valine [d]	3.0	Up
16.44[a]	132.1019	0.1	(M+H)	$C_6H_{13}NO_2$	Leucine [d]	5.7	Up
16.67[a]	132.1024	3.7	(M+H)	$C_6H_{13}NO_2$	Isoleucine	6.1	Up
18.46[a]	166.0856	− 4.5	(M+H)	$C_9H_{11}NO_2$	Phenylalanine [d]	5.4	Up
18.87[a]	182.0802	− 5.3	(M+H)	$C_9H_{11}NO_3$	Tyrosine [d]	5.3	Up
20.05[a]	427.0938	− 3.2	(M+H)	$C_{13}H_{22}N_4O_8S_2$	S-Glutathionyl-L-Cysteine	2.0	Up
20.98[a]	307.0839	2.1	(M+2H)	$C_{20}H_{32}N_6O_{12}S_2$	Glutathione, oxidized	3.0	Up
23.21[a]	308.0907	− 1.3	(M+H)	$C_{10}H_{17}N_3O_6S$	Glutathione, reduced	17.0	Up
0.27[b]	146.1654	1.1	(M+H)	$C_7H_{19}N_3$	Spermidine [d]	1.8	Up
0.28[b]	277.1369	− 9.1	(M+H)	$C_{11}H_{20}N_2O_6$	Saccharopine	−1.3	Down
0.29[b]	203.2222	− 3.9	(M+H)	$C_{10}H_{26}N_4$	Spermine [d]	−3.8	Down
0.37[b]	132.0761	− 5.0	(M+H)	$C_4H_9N_3O_2$	Creatine	2.1	Up
0.37[b]	204.1228	− 0.9	(M+H)	$C_9H_{17}NO_4$	Acetylcarnitine [d]	[f]	Up
0.38[b]	258.1102	6.9	(M+H)	$C_{10}H_{15}N_3O_5$	5-Methylcytidine	1.9	Up
0.39[b]	232.1291	− 0.6	(M+H)	$C_9H_{17}N_3O_4$	Val,Gly,Gly; Ala,Ala,Ala; Asn,Val [e]	−1.1	Down
0.41[b]	202.1799	− 1.5	(M+H)	$C_{11}H_{23}NO_2$	11-Amino-undecanoic acid	1.1	Up
0.46[b]	182.0809	− 1.5	(M+H)	$C_9H_{11}NO_3$	Tyrosine	[f]	Up
0.47[b]	307.0836	1.0	(M+2H)	$C_{20}H_{32}N_6O_{12}S_2$	Glutathione, oxidized [d]	2.1	Up
0.52[b]	132.1022	2.4	(M+H)	$C_6H_{13}NO_2$	Leucine [d]	4.2	Up
0.54[b]	268.1041	0.2	(M+H)	$C_{10}H_{13}N_5O_4$	Adenosine [d]	2.3	Up
0.83[b]	166.0862	− 0.4	(M+H)	$C_9H_{11}NO_2$	Phenylalanine [d]	4.0	Up
1.19[b]	253.1185	0.7	(M+H)	$C_{12}H_{16}N_2O_4$	Tyr, Ala [e]	[f]	Up
1.24[b]	220.1177	− 1.0	(M+H)	$C_9H_{17}NO_5$	Pantothenic acid [d]	[g]	Down
1.31[b]	451.1393	1.2	(M+H)	$C_{25}H_{22}O_8$	Irigenin	1.6	Up
1.40[b]	205.0970	− 1.1	(M+H)	$C_{11}H_{12}N_2O_2$	Tryptophan [d]	3.5	Up
2.65[b]	253.0481	− 5.4	(M+2H)	$C_{30}H_{16}O_8$	Hypericin	1.6	Up
3.28[b]	319.1727	0.9	(M+H)	$C_{11}H_{22}N_6O_5$	Ser, Gly, Arg [e]	[f]	Up
3.30[b]	394.1859	6.5	(M+H)	$C_{16}H_{23}N_7O_5$	His, Thr, His [e]	1.1	Up
3.35[b]	264.0523	− 4.9	(M+H)	$C_9H_{13}NO_6S$	Epinephrine sulfate	1.1	Up
4.07[b]	341.1082	1.0	(M+H)	$C_{12}H_{20}O_{11}$	3-Ketolactose	1.1	Up
4.55[b]	555.2984	4.0	(M+Na)	$C_{27}H_{48}O_8S$	5-beta-cyprinolsulfate	[g]	Down
4.65[b]	372.2596	− 2.5	(M+H)	$C_{17}H_{33}N_5O_4$	Lys, Pro, Lys [e]	−1.2	Down
4.92[b]	335.1377	− 2.0	(M+H)	$C_{12}H_{22}N_4O_5S$	Gln, Met, Gly [e]	[g]	Down
5.45[b]	467.3341	− 5.6	(M+H)	$C_{27}H_{46}O_6$	Varanic acid	−1.4	Down
5.74[b]	175.1082	3.0	(M+2H)	$C_{14}H_{28}N_4O_6$	Thr, Lys, Thr [e]	1.1	Up
6.01[b]	492.2161	6.4	(M+H)	$C_{27}H_{29}N_3O_6$	Tyr, Phe, Tyr [e]	−1.2	Down
6.17[b]	218.0482	0.0	(M+H)	$C_8H_{11}NO_4S$	Tyramine-O-sulfate	−1.1	Down
6.31[b]	383.2152	0.6	(M+H)	$C_{15}H_{26}N_8O_4$	Arg, Ala, His [e]	−1.2	Down
6.59[b]	216.1596	1.1	(M+H)	$C_{11}H_{21}NO_3$	N-Nonanoylglycine	[f]	Up
6.86[b]	392.2156	− 6.1	(M+H)	$C_{20}H_{29}N_3O_5$	Tyr, Leu, Pro [e]	1.1	Up
0.67[c]	347.2305	4.6	(M+H)	$C_{15}H_{30}N_4O_5$	Lys, Ser, Ile [e]	1.9	Up

88

Table 1. Continued

Migration/ retention time	Measured m/z	Error (ppm)	Ion	Formula	Tentative identification	Fold change	Regulation
0.69[c]	388.2551	− 0.8	(M+H)	$C_{17}H_{33}N_5O_5$	Lys, Gln, Leu[e]	−12.1	Down
0.83[c]	176.0919	1.1	(M+H)	$C_7H_{13}NO_4$	N-carboxyethyl-gamma-aminobutyric acid	−1.2	Down
1.03[c]	137.0458	0.0	(M+H)	$C_5H_4N_4O$	Hypoxanthine	1.4	UP
1.10[c]	424.1848	5.0	(M+H)	$C_{18}H_{25}N_5O_7$	Tyr, Gln, Asn[e]	1.2	UP
1.24[c]	786.5335	6.9	(M+H)	$C_{42}H_{76}NO_{10}P$	Phosphatidylserine	1.6	UP
1.24[c]	686.4724	− 4.6	(M+H)	$C_{37}H_{68}NO_8P$	Phosphatidylethanolamine	1.3	UP
1.28[c]	136.0618	0.2	(M+H)	$C_5H_5N_5$	Adenine	4.6	UP
2.42[c]	356.1266	− 2.5	(M+H)	$C_{15}H_{21}N_3O_5S$	Ser, Phe, Cys[e]	−4.3	Down
2.56[c]	341.1082	1.0	(M+H)	$C_{12}H_{20}O_{11}$	3-Ketolactose	1.2	UP
2.75[c]	343.1235	− 0.1	(M+H)	$C_{12}H_{22}O_{11}$	Lactose, isolactose, maltose, cellobiose, etc.	1.1	UP
4.61[c]	202.1806	2.1	(M+H)	$C_{11}H_{23}NO_2$	11-Aminoundecanoic acid	1.3	UP

a) Identified metabolites from CE-ESI-TOF analysis.
b) Identified metabolites from RP/UPLC-ESI-TOF analysis.
c) Identified metabolites from HILIC/UPLC-ESI-TOF analysis.
d) Identification confirmed with commercial standards.
e) Dipeptide/tripeptide.
f) Metabolites only detected in polyphenol-treated cells.
g) Metabolites only detected in control cells.

the robustness of the CE-MS system, although a reduction in sensitivity due to the dilution of the sample at the interface by the sheath liquid was unavoidable.

Higher injection volumes can be used in UPLC-MS resulting in a better sensitivity when compared to CE-MS analysis. Higher sample volumes could, however, be a drawback when sample volume is the limiting constraint (particularly when certain biological samples are under study), being CE-MS especially useful in these cases. As can be observed in Table 1, metabolome coverage was improved by the combination of the different analytical procedures. Some of the significantly different metabolites identified were observed in both CE-MS and RP/UPLC-MS (namely, creatine, leucine, phenylalanine, tyrosine, and oxidized glutathione (GSSG)). Good agreement was also observed in terms of fold change using the different platforms as can be deduced from the values given in Table 1. In our hands, HILIC/UPLC-MS was the least informative platform in terms of the number of metabolites tentatively identified, as can be seen in Table 1.

Due to the well-known metabolome complexity, an insufficient resolution of compounds during separation results in a negative effect mostly on low-abundance metabolites due to ion suppression during ESI. Besides, the changing solvent composition in UPLC-MS during the chromatographic separation under gradient condition increases the difficulty for metabolite identification due to different adducts formation along the chromatogram. On the contrary, when using CE-MS, the background composition does not change during the entire separation, reducing the difficulty for metabolite identification. CE-MS and UPLC-MS also provided different variation of migration/retention times between injections. Migration/retention time shifts were less visible in UPLC-MS compared to CE-MS (see Fig. 1 in Supporting Information). Indeed, percent RSD values of migration time in CE-MS were

lower than 3.5% in the intraday ($n = 5$ injections) and 9.4% in the interday (3 days, $n = 15$ injections) for ten metabolites arbitrarily selected along the electropherogram. Likewise, percent RSD values for retention time were lower than 0.6% in the intraday and 0.9% in the interday in RP/UPLC-MS, while percent RSD values were lower than 0.5% in the intraday and 0.8% in the interday for HILC/UPLC-MS analysis. Larger variation in migration time between runs and the wider variety of peak shapes obtained in CE is mainly due to changes in the inner capillary wall induced by the sample matrix complexity. Actually, the poor migration time reproducibility with CE forced the application of the alignment algorithms obtaining higher proportion of false positives that can only be detected by visual examination of extracted ion electropherograms obtained after data processing.

4.3 From metabolites to pathways

From the total number of tentatively identified metabolites and fold change values uploaded into the IPA software, as described in section 2.11, 33 metabolites were mapped to biological functions and pathways. Initial examination of IPA results revealed 22 highly related metabolites, namely, adenine, adenosine, creatine, reduced glutathione (GSH), GSSG, hypoxanthine, alanine, arginine, histidine, isoleucine, lysine, phenylalanine, tryptophan, tyrosine, N-acetylspermidine, phosphatidylethanolamine, phosphatidylserine, S-adenosyl-L-homocysteine, saccharopine, sarcosine, spermidine and spermine. The change in the intracellular concentration of some of these metabolites is essential for the maintenance of the cellular functions. For instance, rosemary polyphenols increased both reduced (GSH) and oxidized (GSSG) glutathione in HT29 cells. Furthermore, the accumulation
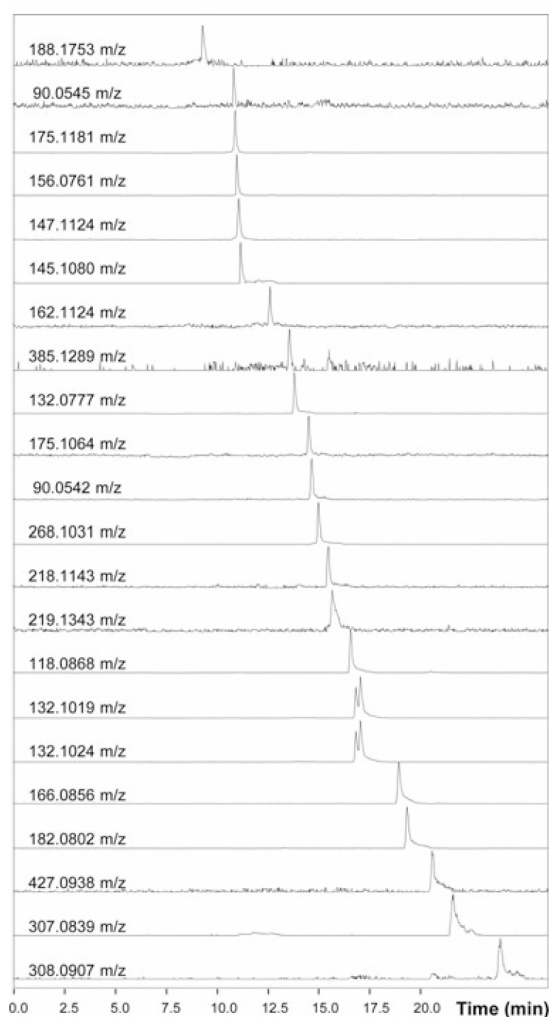
Figure 1. CE-ESI-TOF MS extracted ion electropherograms of significant metabolites identified in polyphenol-treated colon cancer cells.

rate was higher for GSH (fold change of 17) than for the oxidized form (fold change of 3) upon exposure to rosemary polyphenols. Dietary polyphenols have been associated with antioxidant activity, preventing injury caused by free radicals through different mechanisms, including direct radical scavenging, chelating divalent cations involved in Fenton reaction, and modulation of enzymes related to oxidative stress [17]. In this context, glutathione has a central role in the maintenance of the thiol-disulfide redox state in mammalian cells. In this work, the enhanced GSH/GSSG ratio observed in treated cells is in agreement with the reported chemopreventive activity of catechol-type electrophilic compounds such as carnosol and carnosic acid in neuronal cells, the two major polyphenols in assayed rosemary extract [29, 30]. Six other metabolites, namely sarcosine, arginine, cre-

atine, N-acetyl-L-ornithine, spermidine, and spermine were mapped on *urea cycle and metabolism of amino groups* pathway from IPA Knowledge Base (data not shown). Examination of this pathway indicated changes in the levels of intracellular polyamines, namely, spermine and spermidine, upon exposure to rosemary polyphenols. At physiological pH, these positively charged molecules in addition to putrescine may act as ligands at multiple sites on macromolecules, mainly RNA and DNA. Polyamines are mainly involved in the regulation of gene expression by altering DNA structure and by modulating signal transduction pathways [31]. These metabolites regulate important cellular processes including cell proliferation and viability [32]. Hence, intracellular polyamine levels are homeostatically maintained by processes involving biosynthesis, catabolism, and transport [33]. Alterations in polyamine homeostasis lead to changes in intracellular polyamine pools that have important implications for cell growth. In several reports, elevated intracellular polyamines have been associated to certain diseases, including colon cancer [34]. In this work, spermidine accumulated (fold change of 1.8) more in cells incubated with rosemary polyphenols than in control cells. Coincidently, a significant lower level (fold change of −3.8) of spermine was detected in cells incubated with the extract in comparison with the controls. Accumulation of intracellular spermidine in treated cells may be due to the inhibition of the spermine synthase activity or, alternatively, it could be the result of the increased conversion of spermine to spermidine. The later enzymatic process may progress directly, through the catalytic action of spermine oxidase (SMOX), and/or indirectly through the combined activity of other enzymes, spermidine/spermine N-acetyltransferase (SSAT) and acetylpolyamine oxidase (PAOX) [31]. Interestingly, the levels of N-acetylspermidine, the catalytic product of SSAT enzyme, increased greatly in treated cells (fold change of 6.3). SSAT is the rate-limiting enzyme of polyamine catabolism and an important regulatory step in maintaining polyamine content. This enzyme adds terminal acetyl groups to spermidine and spermine that subsequently promotes the export of acetylated polyamines. Efflux of acetylated polyamines has been shown to occur in combination with the SLC3A2 transporter by a polyamine/arginine exchange reaction promoting intracellular polyamine depletion [35, 36]. However, polyamine efflux is often followed by a compensatory increase in polyamine biosynthesis and presumably to increased flow through the pathway [37]. This increased metabolic pathway has been shown to provoke inhibition of cell growth due to the depletion of pathway precursors, such as ornithine, methionine, adenosyl-L-methionine, and acetyl-CoA, and alternatively, due to the overproduction of toxic pathway byproducts, such as 5′-methylthioadenosine, hydrogen peroxide, and aldehydes [38, 39].

In resume, it has been shown that under the conditions assayed in this work, rosemary polyphenols markedly affected the intracellular levels of polyamines and their derived catabolites on HT29 cells. However, potential identified biomarkers in this early study should be further validated with a larger number of biological replicates to gain acceptance for

widespread use, including the study of some essential topics on polyphenols bioavailability.

Further progress in cancer metabolomics will greatly depend on the improvement of metabolome coverage and bioinformatics platforms. Moreover, taking into account the growing number of metabolomic studies, the creation and adoption by the scientific community of standards for data management, analysis, processing, and reporting is becoming more and more critical.

5 References

[1] Dettmer, K., Hammock, B. D., *Environ. Health Perspect.* 2004, *112*, 396–397.

[2] Simó, C., Ibáñez, C., Gómez-Martínez, A., Ferragut, J. A., Cifuentes, A., *Electrophoresis* 2011, *32*, 1765–1777.

[3] Zhang, A. H., Sun, H., Wang, Z. G., Sun, W. J., Wang, P., Wang, X. J., *Planta Med.* 2010, *76*, 2026–2035.

[4] Kim, H. K., Choi, Y. H., Verpoorte, R., *Trends Biotech.* 2011, *29*, 267–275.

[5] Ramautar, R., Mayboroda, O. A., Somsen, G. W., de Jong, G. J., *Electrophoresis* 2011, *32*, 52–65.

[6] Barbas, C., Moraes, E. P., Villaseñor, A., *J. J. Pharm. Biomed. Anal.* 2011, *55*, 823–831.

[7] Levandi, T., Leon, C., Kaljurand, M., García-Cañas, V., Cifuentes, A., *Anal. Chem.* 2008, *80*, 6329–6335.

[8] León, C., Rodríguez, I., Lucio, M., García-Cañas, V., Ibáñez, E., Schmitt-Kopplin, P., Cifuentes, A., *J. Chromatogr. A* 2009, *1216*, 7314–7323.

[9] Moraes, E. P., Rupérez, F. J., Plaza, M., Herrero, M., Barbas, C., *Electrophoresis* 2011, *32*, 2055–2062.

[10] Sugimoto, M., Hirayama, A., Robert, M., Abe, S., Soga, T., Tomita, M., *Electrophoresis* 2010, *31*, 2311–2318.

[11] Issaq, H. J., Fox, S. D., Chan, K. C., Veenstra, T. D., *J. Sep. Sci.* 2011, *34*, 3484–3492.

[12] Manson, M. M., Gescher, A., Hudson, E. A., Plummer, S. M., Squires, M. S., Pringent, S. A., *Toxicol. Lett.* 2000, *112*, 499–505.

[13] Johnson, I. T., *Proc. Nutr. Soc.* 2007, *66*, 207–215.

[14] García-Cañas, V., Simó, C., León, C., Cifuentes, A., *J. Pharm. Biomed. Anal.* 2010, *51*, 290–304.

[15] Araújo, J. R., Gonçalvesa, P., Martel, F., *Nutr. Res.* 2011, *31*, 77–87.

[16] Fresco, P., Borges, F., Marques, M. P. M., Diniz, C., *Curr. Pharm. Des.* 2010, *16*, 114–134.

[17] Ramos, S., *Mol. Nutr. Food Res.* 2008, *52*, 507–526.

[18] Ngo, S. N. T., Wiliams, D. B., Head, R. J., *Crit. Rev. Food Sci. Nutr.* 2011, *51*, 946–954.

[19] Herrero, M., Simó, C., García-Cañas, V., Ibáñez, E., Cifuentes, A., *Mass. Spec. Rev.* 2012, *31*, 49–69.

[20] Herrero, M., Plaza, M., Cifuentes, A., Ibáñez, E., *J. Chromatogr. A* 2010, *31*, 205–228.

[21] Smith, C. A., Want, E. J., Tong, G. C., Abagyan, R., Siuzdak, G., *Anal. Chem.* 2006, *78*, 779–787.

[22] Tautenhahn, R., Böttcher, C., Neumann, S., *BMC Bioinformatics* 2008, *9*, 504–519.

[23] Prince, J. T., Marcotte, E. M., *Anal. Chem.* 2006, *78*, 6140–6152.

[24] Alonso, A., Julià, A., Beltran, A., Vinaixa, M., Díaz, M., Ibañez, L., Correig, X., Marsal, S., *Bioinformatics* 2011, *27*, 1339–1340.

[25] Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L. Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., Forsythe, I., *Nucleic Acids Res.* 2009, *37*, D603–D610.

[26] Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., *Ther. Drug Monit.* 2005, *27*, 747–751.

[27] Kanehisa, M., Goto, S., *Nucleic Acids Res.* 2000, *28*, 27–30.

[28] Bolton, E., Wang, Y., Thiessen, P. A., Bryant, S. H., *Ann. Rep. Comput. Chem.* 2008, *4*, 217–241.

[29] Satoh, T., Izumi, M., Inukai, Y., Tsutsumi, Y., Nakayama, N., Kosaka, K., Shimojo, Y., Kitajima, C., Itoh, K., Yokoi, T., Shirasawa, T., *Neurosci. Lett.* 2008, *434*, 260–265.

[30] Satoh, T., Kosaka, K., Itoh, K., Kobayashi, A., Yamamoto, M., Shimojo, Y., Kitajima, C., Cui, J., Kamins, J., Okamoto, S., Izumi, M., Shirasawa, T., Lipton, S. A., *J. Neurochem.* 2008, *104*, 1116–1131.

[31] Igarashi, K., Kashiwagi, K., *Int. J. Biochem. Cell Biol.* 2010, *42*, 39–51.

[32] Gerner, E. W., Meyskens, F. L., *Nature Rev.* 2004, *4*, 781–792.

[33] Linsalata, M., Russo, F., *Nutrition* 2008, *24*, 382–389.

[34] Paz, E. A., García-Huidobro, J., Ignatenko, N. A., *Adv. Clin. Chem.* 2011, *54*, 45–70.

[35] Uemura, T., Yerushalmi, H. F., Tsprailis, G., Stringer, D. E., Pastorian, K. E., Hawell, L., Byus, C. V., Gerner, E. W., *J. Biol. Chem.* 2008, *283*, 26428–26435.

[36] Uemura, T., Stringer, D. E., Blohm-Mangone, K. A., Gerner, E. W., *Am J. Physiol. Gastrointest. Liver Physiol.* 2010, *299*, G517–G522.

[37] Kramer, D. L., Diegelman, P., Jell, J., Vujcic, S., Merali, S., Porter, C. W., *J. Biol. Chem.* 2008, *283*, 4241–4251.

[38] Ha, H. C., Woster, P. M., Yager, J. D., Casero, R. A., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 11557–11562.

[39] Vujcic, S., Halmekyto, M., Diegelman, P., Gan, G., Kramer, D. L., Janne, J., Porter, C. W., *J. Biol. Chem.* 2000, *275*, 38319–38328.

SUPPORTING INFORMATION

2. MATERIALS AND METHODS

2.3. Cell culture and growth conditions

Antiproliferative activity assays

To study the effect of rosemary extracts on the proliferation of HT29, cells were seeded onto 60 mm-diameter culture plates at 10000 cells/cm^2, permitted to adhere overnight at 37 ºC and exposed to different rosemary extracts containing 0-10 µM total polyphenols for 0-72 h. After incubation with the rosemary extracts for the indicated time in each case, cell proliferation was estimated by the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) assay [1] as follows: the MTT reagent was added and incubated for 3 hours at 37ºC in humidified 5% CO_2/air atmosphere. After the incubation, the media were aspirated and 200 µL of DMSO were added to each well to dissolve the formazan product by shaking for 30 min. Then, the absorbance at 570 nm was measured in a microplate reader (Anthos Labtec Instruments GmbH, Wals, Austria). Rosemary extracts concentrations of 0.1 and 1 µM of polyphenols did not exert significant effect on colon cancer cell proliferation after 72 h incubation. Incubation with higher concentrations (10 µM polyphenols), reduced cell proliferation after treatment for 48h (data not shown).

REFERENCES

[1] Mosmann, T., *J. Immunol. Methods* 1983, *65,* 55-63.

Supporting Figures
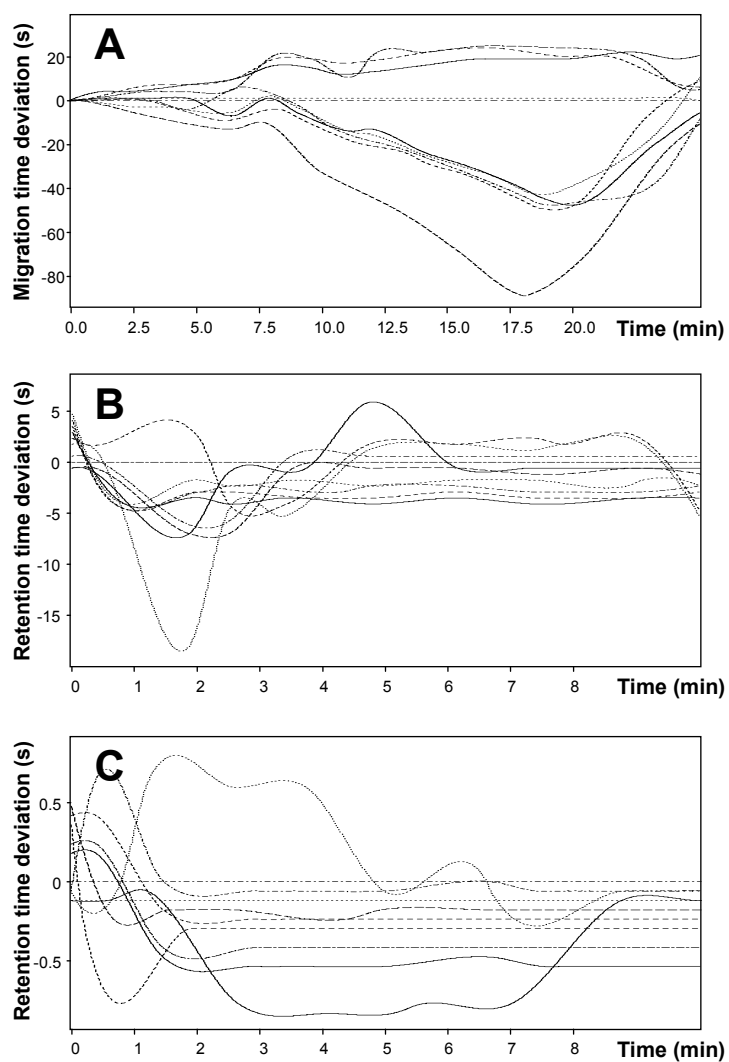
Figure 1-S. Analysis time variations obtained with the three analytical platforms. (A) CE-MS, (B) C18/UPLC-MS and (C) HILIC/UPLC-MS.

3.4. CAPÍTULO 3 (Chapter 3)

Global foodomics strategy to investigate the health benefits of dietary constituents

Global Foodomics strategy to investigate the health benefits of dietary constituents

Clara Ibáñez[a], Alberto Valdés[a], Virginia García-Cañas[a], Carolina Simó[a], Mustafa Celebier[a], Lourdes Rocamora-Reverte[b], Ángeles Gómez-Martínez[b], Miguel Herrero[a], María Castro-Puyana[a], Antonio Segura-Carretero[c], Elena Ibáñez[a], José A. Ferragut[b], Alejandro Cifuentes[a,*]

[a] Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain
[b] Institute of Molecular and Cellular Biology, Miguel Hernandez University, Avda. Universidad s/n, Elche, Alicante, Spain
[c] Faculty of Sciences, University of Granada, Fuentenueva s/n, 18071 Granada, Spain

ARTICLE INFO

ABSTRACT

A global methodology, called Foodomics, which allows carrying out a comprehensive evaluation of the health benefits of food ingredients is presented in this work. The new methodology is based on the combination of several analytical platforms and data processing for Transcriptomics, Proteomics and Metabolomics studies, allowing the determination of changes induced by food ingredients at molecular level. Both, the whole methodological development and its potential are presented through the investigation of a case study following a hypothesis-free strategy. Namely, the chemopreventive effect of polyphenols from rosemary was examined on the total gene, protein and metabolite expression in human HT29 colon cancer cells. Conclusions on the bioactivity of polyphenols against colon cancer cells based on the results from each single platform (Transcriptomics, Proteomics or Metabolomics) are compared with the conclusions based on the integration of the whole results from the three platforms, corroborating the interest of using a global integrative strategy as Foodomics. To our knowledge, although many papers and reviews have been published on this topic, this is the first time that Transcriptomics, Proteomics and Metabolomics platforms are put together to study the health benefits from dietary ingredients against colon cancer cells at gene, protein and metabolite level. Advantages, drawbacks and current challenges of this global analytical strategy are discussed in this work. The results from our study provide new insights on the biological mechanisms involved in the cancer risk reduction properties of dietary constituents.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, the general shift observed in medicine and biosciences toward prevention of future diseases has brought about the generation of a new field of research in Food Science and Nutrition in which food is investigated not only as source of energy but also as potential health promoter. As a result, food scientists and nutritionists working on this new field of research are facing a large number of new challenges to answer the demands from consumers, agencies, regulatory laboratories and food industries [1]. One of the main challenges is to improve our limited understanding of the roles of nutritional compounds at molecular level (i.e., their interaction with genes and their subsequent effect on proteins and metabolites) for the rational design of strategies to manipulate cell functions through diet, which is expected to have an extraordinary impact on our health [2,3]. This trend has given rise to the development of new methodologies in which advanced analytical methodologies, mainly "omics", and bioinformatics – frequently together with in vitro, in vivo and/or clinical assays – are applied to investigate topics considered unapproachable few years ago.

Foodomics has been recently defined by our group as a discipline that studies the Food and Nutrition domains through the application of advanced omics technologies to improve consumer's well-being, health, and confidence [4–6]. Thus, the huge analytical potential of *Foodomics* can allow solving questions related to food safety, traceability, quality, new foods, transgenic foods, functional foods, nutraceuticals, etc. [4–6] *Foodomics* can therefore be an adequate strategy to investigate the complex issues related to prevention of future diseases through food intake. In this regard, it is interesting to remark that several of the health benefits assigned to many dietary constituents [7–11] are still under controversy as can be deduced from the large number of applications rejected by

the European Food Safety Authority about health claims of new foods and ingredients [12]. More sounded scientific evidences are needed to demonstrate or not the claimed beneficial effects of these new foods and constituents. In this sense, the advent of new post-genomic strategies as *Foodomics* seems to be essential to understand how the bioactive compounds from diet interact at molecular and cellular level, as well as to provide better scientific evidences on their health benefits. In spite of this general need, it is practically impossible to find in the literature works in which several omics techniques are integrated to investigate the health benefits of a dietary ingredient (or group of ingredients) at gene, protein and metabolite level, simultaneously [13]. The combination of the information from the three expression levels can be crucial to adequately understand and scientifically sustain the health benefits from food ingredients. For instance, many works have been published so far studying the effect of dietary polyphenols on different types of cancer (for an overview on this topic see, e.g., the review works [7,9,14]). However, it has been repeatedly indicated that, effects from dietary polyphenols involve multiple molecular and biochemical mechanisms of action, which are still not completely characterized [7], concluding that many features remain to be elucidated about their claimed activity [15].

In 2008, colorectal cancer was the third most common cancer worldwide in men (663,000 cases, 10.0% of the total) and the second in women (571,000 cases, 9.4% of the total). About 608,000 deaths from colorectal cancer were estimated worldwide, accounting for 8% of all cancer deaths, making it the fourth most common cause of death from cancer, with the highest mortality rates estimated in Central and Eastern Europe (20.1 per 100,000 for male, 12.2 per 100,000 for female) [16]. Although surgical resection is an effective treatment for localized disease achieving a 5-year survival rate of 90%, other normal treatments for metastatic disease remain ineffective. It is known that Mediterranean countries have lower colorectal cancer rates than other Western countries and it is now generally assumed the important role that diet plays in this illness [17,18]. Vegetable consumption is thought to protect against colorectal cancer, while high fat and meat intakes are known risk factors. It has been also mentioned, that inhibition of proliferation constitutes an interesting target for chemoprevention in which intervention could commence before the period of tumor proliferation preventing the series of events leading to metastasis [19]. In relation to the poor prognosis of colorectal cancer [20] a window of opportunity exists for a chemopreventive regimen to be implemented in line with other therapies. However, little is known about the changes at the genome/proteome/metabolome levels that may be related to or involved in this protective mechanism.

In the present work, a new Foodomics platform is presented that allows carrying out a global evaluation on the health benefits of dietary constituents at different expression levels. To demonstrate the potential of this strategy, the chemopreventive effect of dietary polyphenols against human colon cancer cells was selected as case study. The new Foodomics platform combines different analytical techniques and data processing for Transcriptomics, Proteomics and Metabolomics studies, allowing the determination of genes, proteins and metabolites that show statistically different expression. This global *Foodomics* platform is presented together with their specific data processing tools as an effective strategy to characterize changes at the transcriptome, proteome and metabolome level. This study follows a hypothesis-free approach in which the biological information from each expression level considered individually is compared with the biological information obtained after combining the three expression levels in a global *Foodomics* strategy. Advantages, drawbacks and current challenges of this global strategy are discussed in this work. *Foodomics* is shown to provide added biological information on the mechanisms involved in the cancer risk reduction properties of dietary polyphenols.

2. Materials and methods

2.1. Chemicals

All chemicals were of analytical reagent grade. A phosphate buffered saline (PBS) solution (138 mM sodium chloride, 2.7 mM potassium chloride and 10 mM sodium hydrogen phosphate, at pH 7.4) and a protease inhibitor cocktail (containing 4-(2-aminoethyl)benzenesulfonyl fluoride, pepstatin A, E-64, bestatin, leupeptin and aprotinin) for HT29 cell culture, were purchased from Sigma–Aldrich (St. Louis, MO, USA). For Proteomics, precision plus protein unstained standard used as molecular weight marker on SDS-PAGE and SYPRO Ruby Protein Gel Stain used for staining the gels were from Bio-Rad (Hercules, CA, USA); trichloroacetic acid (TCA), DL-dithiothreitol (DTT), iodoacetamide (IAA), urea, SDS, tris(hydroxymethyl)aminomethane (Tris), glycerol, ethylenediaminetetraacetic acid (EDTA), glycine, bromophenol and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were from Sigma–Aldrich; HCl for pH adjustment was from Probus (Barcelona, Spain); and finally, acetic acid, methanol and acetone were from Scharlau (Barcelona, Spain). Reagents and solvents employed in Metabolomics for the preparation of LC mobile phases, CE electrolytes, sheath liquid and standard metabolite solutions were of MS grade: formic acid and 2-propanol were from Riedel-de Haën (Seelze, Germany), acetonitrile and water were from Labscan (Gliwice, Poland). Metabolite standards were also purchased from Sigma–Aldrich.

2.2. Polyphenols extract

Extracts rich in dietary polyphenols were obtained from rosemary (*Rosmarinus officinalis*) using supercritical fluid extraction (SFE) under the following conditions of pressure and solvent: supercritical CO_2 plus 7% (v/v) ethanol at 150 bar and 40 °C for 300 min were used as previously reported by Herrero et al. [21]. Chemical characterization of this extract provided information on the major phenolic constituents, namely cirsimaritin, genkwakin, carnosol, carnosic acid and methylcarnosate.

2.3. Colon cancer cell culture and growth conditions

Human colon cancer HT29 cells were grown in Dulbecco's modified eagle medium (DMEM) supplemented with 5% (v/v) heat-inactivated fetal calf serum, 2 mM L-glutamine, 50 μg/mL penicillin G and 50 μg/mL streptomycin, at 37 °C in humidified atmosphere and 5% (v/v) CO_2. Cells were plated at a density of 10000 cells/cm^2 in 60-mm diameter culture plates and exposed for 72 h to a dissolution containing 10 μM of total rosemary phenols (i.e., treated cells). The respective control cells were grown under identical conditions but without the rosemary phenols (i.e., control cells).

2.4. Cell cycle study

To determine whether rosemary polyphenols regulate cell cycle progression of colon cancer cells in vitro, HT29 cells were treated with the above mentioned extract, and DNA was stained with propidium iodide followed by flow cytometry. After 72 h of treatment, accumulation of HT29 cells in G2/M cell cycle phase was observed. This increase of cells in G2/M phase was accompanied by a corresponding reduction in the proportion of cells in G1 phase. Moreover, the assayed rosemary polyphenols gave rise to a modest pro-apoptotic effect supported by a slight accumulation of subG1 populations of HT29 cells. These data indicate that these polyphenols promote G2/M cell cycle arrest and apoptosis in HT-29 colon cancer cells.

2.5. Transcriptomics

Total RNA was isolated from triplicate samples of treated and control HT29 colon cancer cells using RNeasy Mini Kit (Qiagen, Spain) according to manufacturer protocol. The quality of the isolated RNA was determined with a NanoDrop ND1000 (Thermo, Spain) and a Bioanalyzer 2100 from Agilent Technologies (Palo Alto, CA, USA). Sample preparation and microarray hybridization onto Human Gene 1.0ST chips was carried out by Servicio de Genómica (Parque Científico de Madrid, Spain). Quality assessment of raw data (CEL files) and pre-processing were performed using Expression Console™ (Affymetrix) and Robust Multi-Array (RMA) normalization in the BioConductor package affy for R, respectively [22,23]. Significance analysis and multiple testing correction [24] were performed using the BioConductor package limma [25] in order to control or estimate the false discovery rate in the datasets.

To identify the differentially expressed genes (DEGs), microarray data was subjected to gene filtering based on a combination of M-value cutoff, which represents a $\log_2$-fold change between the two experimental conditions (treatment with polyphenols vs. control), and the statistical significance (false discovery rate, FDR applied on moderated t-statistics). In this study, DEGs were identified based on 0.7 as M-value cutoff that corresponds to expression ratios (fold-change) ≥ 1.6 for up-regulated and ≤ 0.6 for down-regulated genes; and the statistical filter was established at 5% FDR (adjusted p-value <0.05).

The confirmation of microarray results was carried out by RT-qPCR. To this aim, a selection of genes was made based on their degree of expression change, adjusted p-value, and/or known biological function. Starting amounts of 0.5 μg of total RNA isolated from cells were reverse transcribed in a volume of 20 μL using Transcriptor First Strand cDNA Synthesis kit with oligo(dT) primers (Roche Diagnostics, Barcelona, Spain). Each real-time quantitative PCR reaction was performed on 0.5 μL aliquots of diluted (1:10) cDNA solutions using LightCycler® 480 Real-Time PCR and LightCycler® 480 Probes Master kit (Roche Diagnostics). Target-specific PCR primers and Human Universal Probe Library probes were selected using the Probe Finder assay design software (Roche Diagnostics, http://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp). Primers were designed to have melting temperature values close to 60 °C and to span exon–exon junctions. The designed primers were then checked with Oligo Analyzer 3.1 software (Integrated DNA Technologies, http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer) to predict homodimers, heterodimers and possible secondary structures, and to redesign the primers if needed. The primers were purchased from Fisher Scientific (Alcobendas, Spain). Two technical replicates were performed for each sample in a 96-well format plate. On each plate in parallel, four endogenous control genes (*GAPDH*, *B2M*, *IPO8* and *PPIA*) and no-template-controls (NTC) were also performed in duplicate. Second derivative maximum method [26] was used to calculate cycle-threshold (Ct) values. All primers utilized displayed PCR efficiencies higher than 90%. Using the Relative Expression Software Tool (REST) the relative expression of selected genes was calculated using efficiency correction option, and compared among treated and control cells. The randomization test method, as a part of the REST software, was used to assess statistical significance in terms of up- or down-regulation of target genes after normalization to the four reference genes.

Functional enrichment analysis was performed as a previous step for a reliable data interpretation obtained from microarray analysis. To this aim, two computational tools were used, namely Ingenuity Pathway Analysis (IPA; Ingenuity Systems, USA) and Gene Set Enrichment Analysis (GSEA) [27,28]. First, IPA software was employed to interpret the gene expression data in the context of biological processes and pathways. Based on the list of up- and down-regulated identifiers, IPA performs functional enrichment analysis in order to identify the biological processes and functions over-represented in a given list of genes. Significance of the Molecular and Cellular Functions, as well as the Signaling Pathways was tested by the Fisher Exact test p-value. Exploratory functional analysis was also performed with GSEA as an alternative approach to identify biological processes and functions that are modulated upon exposure to rosemary polyphenols. GSEA algorithm ranked all microarray genes according to their expression under each experimental condition (treated and control). This enrichment strategy uses a priori defined gene sets to determine whether the members of a given gene set are primarily found at the top (induced gene expression) or bottom (repressed gene expression) of the rank list of genes, or to the contrary, are randomly distributed throughout the ranking. To this end, normalized enrichment scores (NES) are calculated in order to determine the extent to which individual genes from a gene set are represented at the extremes of the ranked gene list. In this study, GSEA analyses were conducted using the C5.BP catalog of gene sets from Molecular Signatures Database v3.0 (MSigDB, http://www.broadinstitute.org/gsea/msigdb/index.jsp). FDR% and significant p-values, calculated relative to NES values, were defined as 25% and less than 0.05, respectively.

2.6. Proteomics

Human HT29 colon cancer cells were washed with PBS solution and centrifuged. The pellet was resuspended with homogenization buffer (10 mM Tris–HCl, 5 mM EDTA, 120 mM NaCl, at pH 7.4) and protease inhibitor cocktail. Cells were disrupted with a Polytron homogenizer and centrifuged (14 min at 14,000× g and 4 °C). Pellet (mainly containing nuclear and mitochondrial fractions) was discarded and supernatant was centrifuged for 1 h at 100,000 × g and 4 °C. Supernatant (cytosolic fraction) was stored at −80 °C until analysis. The total protein amount from the cytosolic fraction was determined by using DC protein assay from Bio-Rad. Protein precipitation was carried out with 10% (v/v) TCA and 20 mM DTT in acetone at −20 °C. Protein pellets were dissolved in the "2-DE sample buffer" (8 M urea, 2% (m/v) CHAPS, 50 mM DTT, 0.2% (v/v) Bio-Lyte 3/10 ampholyte, 0.001% (m/v) bromophenol) to a final protein concentration of 1 mg/mL. Each 11-cm long IPG pH 5–8 strip was rehydrated with 280 μL of protein solution for 5 hrs. After isoelectric focusing (IEF) the IPG strips were equilibrated for 10 min in equilibration buffer 1 (6 M urea, 2% (m/v) SDS, 0.375 M Tris–HCl at pH 8.8, 20% (m/v) glycerol and 2% (m/v) DTT) and for 10 min in equilibration buffer 2 (6 M urea, 2% (m/v) SDS, 0.375 M Tris–HCl at pH 8.8 and 20% (m/v) glycerol) under shaking. After equilibration, the IPG strips were laid on a 4–12% (m/v) Bis–Tris Criterion XT Precast Gel from Bio-rad with 0.5% (m/v) agarose in the cathode buffer (192 mM glycine, 0.1% (m/v) SDS and Tris–HCl at pH 8.3). SYPRO Ruby (from Bio-rad) was used for gel staining. The 2-DE gels were scanned with a Versa-Doc image system (Bio-Rad). The relative gels images were analyzed by using the PDQuest software (Bio-Rad). Each sample was analyzed by triplicated.

Spots of interest were manually excised and the digested with trypsin [29]. For MALDI-TOF/TOF-MS analysis, samples were automatically acquired in an ABI 4800 MALDI-TOF/TOF-MS (Applied Biosystems, Framingham, MA, USA) in positive ion reflector mode (the ion acceleration voltage was 25 kV to MS acquisition and 1 kV to MS/MS) and the obtained spectra were stored into the ABi 4000 Series Explorer Spot Set Manager. To submit the combined peptide mass fingerprinting (PMF) and MS/MS data to MASCOT software v.2.2.04 (Matrix Science, London, UK), GPS Explorer v4.9 was used, searching in the non-redundant NCBI protein database (NCBI no. 20100930 (11,960,556 sequences; 4,082,908,561 residues)).

The confidence interval for protein identification was set to ≥95% ($p < 0.05$) and only peptides with an individual ion score above the identity threshold were considered as correctly identified.

2.7. Metabolomics

The cytosolic fraction from control and polyphenol-treated HT29 colon cancer cells was obtained as described in Section 2.6. Metabolite purification was necessary considering the reasonably high amount of proteins previously measured in the cytosolic fraction. For this purpose, 400 μL of the isolated cytosol obtained from both treated and control cell cultures, were ultrafiltrated using Amicon Ultra 3 kDa centrifugal devices from Millipore (Bedford, MA, USA) at 14,000 × g and 20 °C for 40 min. Low molecular weight fraction (less than 3 kDa) was aliquot and stored at −80 °C until CE-MS or UPLC–MS analysis. In the present metabolomic platform all samples were analyzed by quintuplicate.

Metabolomic analyses with capillary electrophoresis (CE) were carried out in a P/ACE 5500 CE apparatus from Beckman Instruments (Fullerton, CA, USA). The instrument was controlled by a PC running the System Gold software from Beckman. CE was coupled to MS through an orthogonal electrospray ionization (ESI) interface G1607A model from Agilent Technologies. Electrical contact at the electrospray needle tip was established via a sheath liquid delivered by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). A TOF-MS instrument (micrOTOF model) from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a PC running the micrOTOF control software from Bruker Daltonics. New fused-silica capillaries from Composite Metal Services (Worcester, England) were conditioned by rinsing with 1 M NaOH for 10 min followed by 20 min with water both using pressurized N$_2$ at 20 psi (1380 mbar). Between runs, capillaries were flushed with water during 2 min, followed by separation buffer during 4 min. Injections were made at the anodic end for 80 s using a N$_2$ pressure of 0.5 psi (34.5 mbar). For CE-MS analysis, 1 M formic acid was used as background electrolyte (BGE). The sheath liquid consisted of isopropanol–water (50:50, v/v) and was delivered at a flow rate of 0.24 mL/min. Separation voltage was +25 kV and capillary temperature was set at 25 °C. The mass spectrometer operated with the ESI source in the positive ion mode. The nebulizer and drying gas conditions were 0.4 bar N$_2$ and 4 L/min N$_2$, respectively, and maintaining the ESI chamber temperature at 250 °C. Data were acquired in the mass range from 50 to 600 m/z. Accurate mass data of the molecular ions were processed using the DataAnalysis 4.0 software from Bruker Daltonics. External and internal calibration of the TOF-MS instrument was performed by introducing a 10 mM sodium formate solution through the separation capillary. The ions used for the calibration of the TOF-MS instrument were next: 90.9766, 158.9641, 226.9515, 294.9389, 362.9263, 430.9138, 498.9012 and 566.8886 m/z.

The ultra-high performance liquid chromatography (UPLC) system Agilent 1290 (Agilent Technologies) was connected to a quadrupole-time-of-flight (Q/TOF) mass spectrometer Agilent 6540 Q/TOF-MS (Agilent Technologies) equipped with an orthogonal electrospray interface with the Jet Stream thermal gradient focusing technology from Agilent and operating in positive ion mode. The instrument was controlled by a PC running the Mass Hunter Workstation software from Agilent. Reversed phase chromatographic separation was performed with an Agilent ZORBAX C18, Rapid Resolution HT (2.1 mm × 50 mm, 1.8 μm) column maintained at 40 °C. Elution was performed using phase A (water with 0.1% (v/v) formic acid) and phase B (acetonitrile with 0.1% (v/v) formic acid) and the following gradient program: the run was started at 2% B and was increased linearly till reached 20% at 6 min, then it was increased linearly till reached 100% at

10 min. After each analysis, 100% B was kept constant for 2 min more and then the column was re-equilibrated for 5 min using the initial solvent composition. Hydrophilic interaction liquid chromatography (HILIC) was performed on a ZORBAX HILIC Plus HT (2.1 mm × 50 mm, 1.8 μm) column maintained at 30 °C. Elution was performed using phase A (water with 10 mM ammonium formate at pH 5) and phase B (acetonitrile) and the following gradient program: the run was started at 90% B and decreased linearly till reached 80% at 7 min, then 50% B at 9 min and 0% B at 10 min. Finally, 0% B was kept constant for 2 min and 90% B was set during 5 min with the initial solvent composition for column re-equilibration. TOF-MS operation parameters were the following: capillary voltage, −4000 V; nebulizer pressure, 40 psi; drying gas flow rate, 10 L/min; gas temperature, 300 °C; skimmer voltage, 45 V; fragmentor voltage was 125 V. Data were acquired in the mass range from 50 to 1000 m/z every 200 ms in positive mode. Internal mass calibration of the instrument was carried out using an AJS ESI source with an automated calibrant delivery system. External calibration of the TOF-MS was carried out using next masses: 118.086255, 322.048121, 622.02896, 922.009798, 1221.990637 and 1521.971475 m/z. The reference compound solution for internal mass calibration of the Q/TOF-MS containing 5 mM of purine ([C$_5$H$_5$N$_4$]$^+$ at m/z 121.050873), and 2.5 mM HP-0921, hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazine ([C$_{18}$H$_{19}$O$_6$N$_3$P$_3$F$_{24}$]$^+$ at m/z 922.009798) was from Agilent.

Quality control (QC) samples were used to monitor the performance of the method and increase the credibility of the data. QC was prepared by combining equal aliquots from each sample and injected regularly throughout the run to monitor the stability of the CE/LC–MS method.

CE-MS data were processed using the XCMS open-source package [30] written in the platform-independent programming language R (www.r-project.com). Data signals obtained from LC–MS were analyzed using MassHunter software 4.0 (MH) and Mass Profiler Professional 2.2 (MPP), both from Agilent. Both CE-MS and LC–MS data processing were carried out using a standard HP computer Z400 (2.8 GHz, 4 GB memory) running Windows 7. Comparative analysis was performed by calculating fold changes (FC) together with their respective Welch t-test to assess the significance of the changes. Each ion was tested for its significance. A two-way parametric Student's t-test was applied to compare the significance of changes in measured ion responses for metabolites in cytosol of the treated cells compared to the control cells. Only probability values $p < 0.05$ were considered significant.

After statistical analysis of CE-MS datasets, significant molecular features were subjected to identification process. Metabolite tentative identification was performed by matching the obtained accurate m/z values and those contained in different free available databases, namely: Human Metabolome Database [31], METLIN [32], KEGG compound (http://www.genome.jp) and PubChem (http://pubchem.ncbi.nlm.nih.gov/) with a mass accuracy window of 10 ppm. When available, commercial standards were used to confirm metabolite identification. Generate-Molecular Formula Editor within Data Anaysis 4.0 (Bruker) and Molecular Formula Generator algorithm within MassHunter (Agilent) software were used to support the agreement between the molecular formula generated by the software and the proposed compound from metabolite database search in terms of mass error (ppm) and isotopic pattern similarity. When isomers existed for a given formula, metabolite identification was sorted giving preference to metabolites from central metabolic pathways in KEGG and number of databases containing each metabolite. Finally, metabolomic data sets were analyzed through the use of IPA software. A data set containing KEGG metabolite identifiers and corresponding fold-change values was uploaded into the IPA application.

100
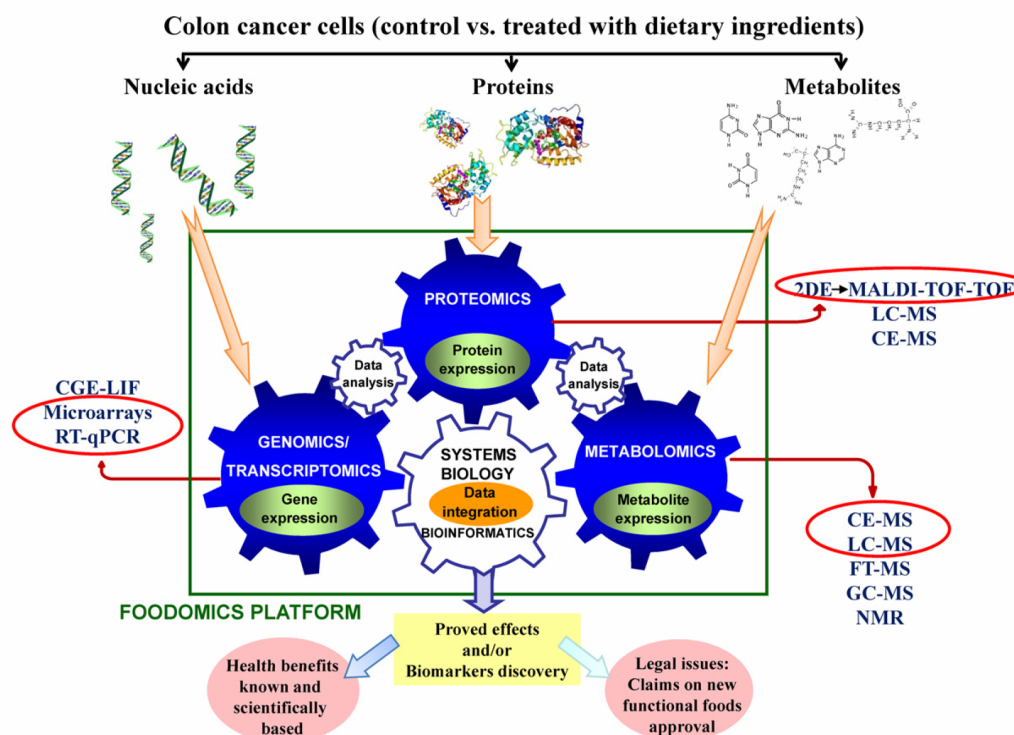
Fig. 1. Scheme of an ideal Foodomics strategy to investigate the health benefits from dietary constituents, including methodologies and expected outcomes. Technologies used in this work are marked with a circle.

2.8. Data integration

In order to link the observed changes across the three omics platforms, an integrative approach based on the networking and mapping capabilities of IPA software was applied. To do this, each metabolite identifier was mapped to its corresponding metabolite object in the Ingenuity Pathway Knowledge Base. Metabolite-centric networks were then algorithmically generated based on their connectivity. Networks were then ranked by a score that defines the probability that a collection of nodes equal to or greater than the number in a network can be achieved by chance alone. According to IPA, a score of 3 indicates that there is a 1/1000 chance that the focus molecules are in a network due to random chance, and therefore, scores >3 have a 99.9% confidence of not being generated by random chance alone. This score was used as the threshold for identifying gene networks that were significantly affected by the treatment with rosemary extract. Then, transcriptomics and proteomics data results were overlaid on top-scored metabolite-centric networks in order to discover associations across the three expression levels.

3. Results

A clear effect at low polyphenols concentration (10 μM) was observed indicating that these dietary polyphenols from rosemary are effective against HT29 colon cancer cells growth and proliferation. Fig. 1 shows the scheme of the ideal *Foodomics* platform in order to investigate the health benefits from dietary constituents, including methodologies and expected outcomes. The technologies used in this work are marked with a circle. Nucleic acids, proteins and metabolites from control and polyphenols-treated HT29 colon cancer cells were obtained as indicated in the experimental

section and analyzed following the global scheme shown in Fig. 2. This figure shows the methodological process followed in this work including a more specific description of the analytical tools developed for Transcriptomics, Proteomics and Metabolomics, as well as the steps for the treatment and processing of data.

3.1. Transcriptomics results

Processing of whole-transcriptome microarrays data [33–35] was done according to the scheme shown in Fig. 3. Following the steps indicated in Fig. 3, a large number of genes were differentially expressed in HT29 colon cancer cell after the polyphenols treatment. As can be seen in the results shown in Fig. 4, considering only the genes whose fold change (FC) was ≤ 0.6 (down-regulated) or ≥ 1.6 (up-regulated), and assuming a p value <0.05 and a false discovery ratio (FDR) of 5%, 1308 genes were found as differentially expressed. According to these Transcriptomics results, rosemary polyphenols altered in the HT29 colon cancer cells the expression of ~4% of the genes covered by the microarray. However, mRNA expression levels often do not reflect the protein and metabolite expression levels, which make the predictive value of mRNA expression levels limited. In the current study, the effect of rosemary polyphenols on protein and metabolite expression in HT29 colon cancer cells were also studied in order to complete the study at the three levels of expression. Among the lists of differentially expressed genes, *HMOX1*, *OSGIN1*, *DUSP1*, *MUC1* and *ARRDC3* genes were selected for RT-qPCR validation, in order to corroborate the validity of the results obtained from microarrays. RT-qPCR expression ratios confirmed the results obtained by microarrays, since a good correlation was observed between data sets obtained by both technologies.
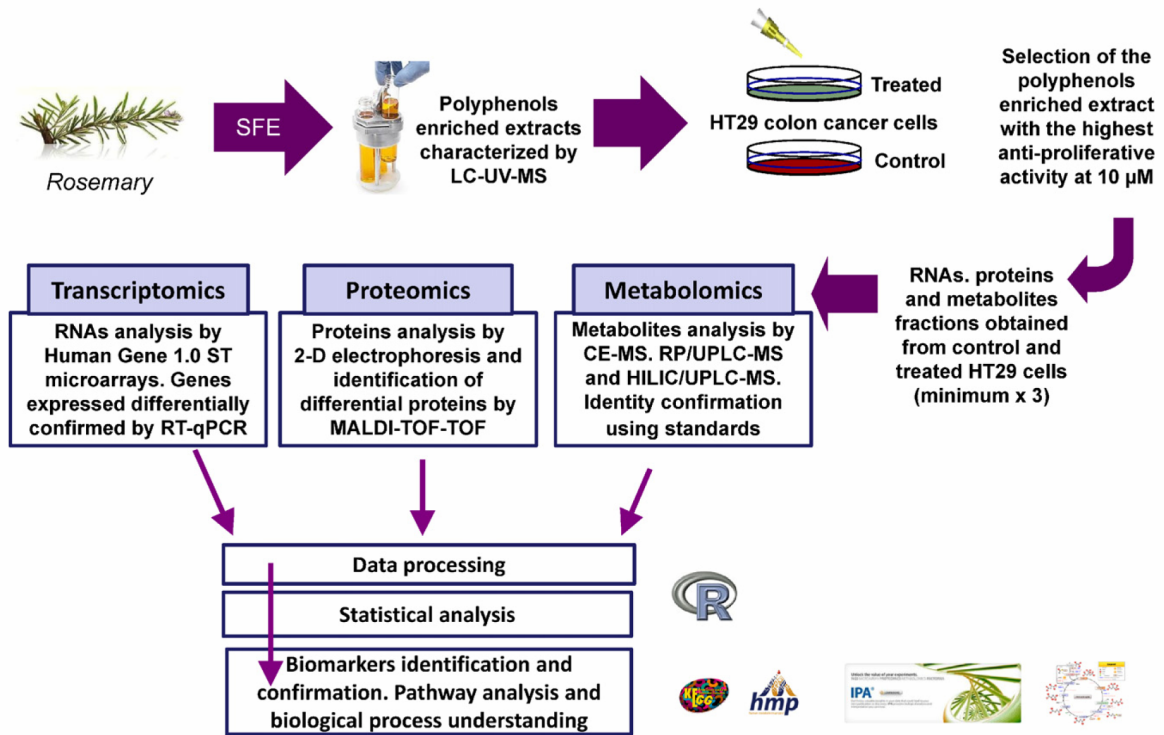
Fig. 2. Global Foodomics strategy used in this work to investigate the activity of rosemary polyphenols against colon cancer HT29 cells at molecular level.
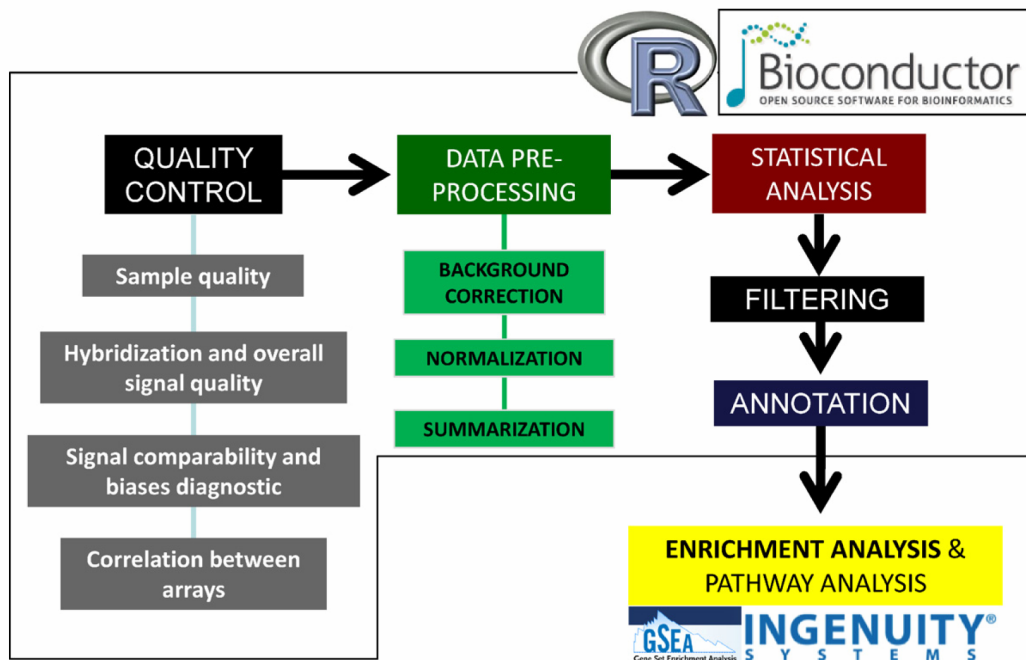
Fig. 3. Scheme of the procedure developed in this work for the processing of data from microarrays based on the open source R programming language. Minimum three biological replicates were used for each experimental condition (i.e., treated vs. control colon cancer cells). A file with extension .cel (raw expression data) is obtained for each replicate from the microarray. The main packages for data processing used in R were: affy, annotate, limma, genefilter, affyQCReport, affyPLM, and ArrayTools.

MA plot

Volcano plot

Setting FDR at 5%
(adjusted p-value <0.05)

0.6 ≥ FC ≥1.6

1308 genes
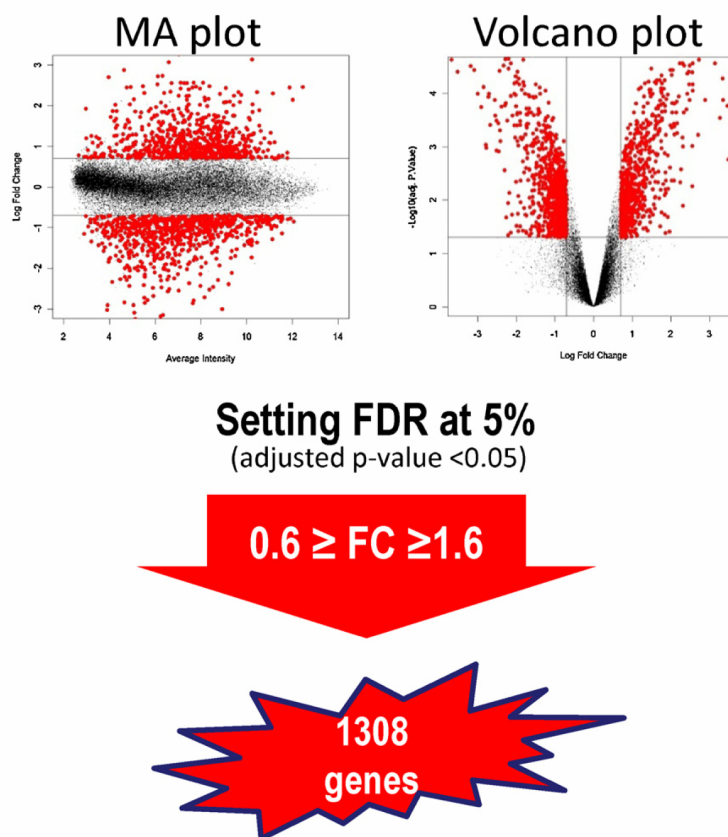
Fig. 4. MA plot (Log Fold Change vs. Average Intensity), volcano plot (Log P value vs. Log Fold Change) and selection of differentially expressed genes (DEGs) after transcriptomics of HT29 colon cancer cells treated with dietary polyphenols. Limma provides a number of summary statistics useful to select DEGs. Moderated t-statistics is used for significance analysis with a p value <0.05. An adjustment method such as false discovery rate (FDR at 5%) is then applied for multiple testing correction. An additional constraint is then included, considering only the genes whose fold change (FC) was lower than 0.6 (down-regulated) or higher than 1.6 (up-regulated). Under these conditions 1308 genes differentially expressed came out.

3.2. Proteomics results

Proteomics was carried out using two-dimensional gel electrophoresis followed by MALDI-TOF/TOF-MS. A scheme of the whole Proteomics process used in this work together with the main results is shown in Fig. 5. Proteins differentially expressed in HT29 colon cancer cells after the treatment with polyphenols were discovered by comparing the proteins fingerprints obtained with the 2-DE gels from control and treated cells. Proteins of interest were excised from the gel and hydrolyzed with trypsin. Analysis of the resulting peptides was done using MALDI-TOF/TOF-MS and proteins next identified using the Mascot database as described in Section 2.6. Proteomics detected 17 proteins that changed their expression in a statistically significant manner. Namely, 10 of these proteins were down-regulated and 7 up-regulated after the treatment with polyphenols as can be seen in Fig. 5.

3.3. Metabolomics results

As indicated in Fig. 2, Metabolomics was carried out using CE-MS, RP/UPLC–MS and HILIC/UPLC–MS, in all cases a TOF-MS was used as mass analyzer. CE is particularly suited for the analysis of ionogenic compounds, while RP/UPLC provides the separation of more hydrophobic compounds. Separation mechanism in HILIC is dominated by polar interactions providing different selectivity than RP/UPLC, while maintaining higher sensitivity and robustness compared to CE. Apart of the different separation mechanism and metabolites coverage, the three techniques also showed some important differences in terms of migration time and retention time repeatability. CE-MS provided the worst repeatability with values up to 60 s in analysis time deviations among injections, while for LC based separations these deviations were much lower (typically up to 5 s). This factor is especially critical in Metabolomics because adequate electropherograms alignment has to be achieved prior to any data processing. Fig. 6A and B show the volcano plots and the number of differentially expressed features after the polyphenols treatment determined by RP/UPLC–MS and HILIC/UPLC–MS, respectively. Using RP/UPLC–MS more than 2176 initial features were detected along the chromatogram, while 2890 features were detected by HILIC/UPLC–MS. These features from control and treated groups were further subjected to data processing, filtering and statistical analysis. After data processing, a total of 210 metabolites showing significant ($p < 0.05$) variation in their abundances as response to the polyphenols treatment were detected by RP/UPLC–MS, while HILIC/UPLC–MS found 214 compounds ($p < 0.05$). CE-MS detected 212 metabolites differently expressed ($p < 0.05$). Metabolome coverage was improved by the combination of the three analytical techniques. An example of this different coverage can be seen in Table 1 that shows a comparison of the 10 metabolites with the higher FC (up-regulated and down-regulated) detected by CE-MS, RP/UPLC–MS and HILIC/UPLC–MS. Only three compounds (phenylalanine, tyrosine and leucine) out of the 30
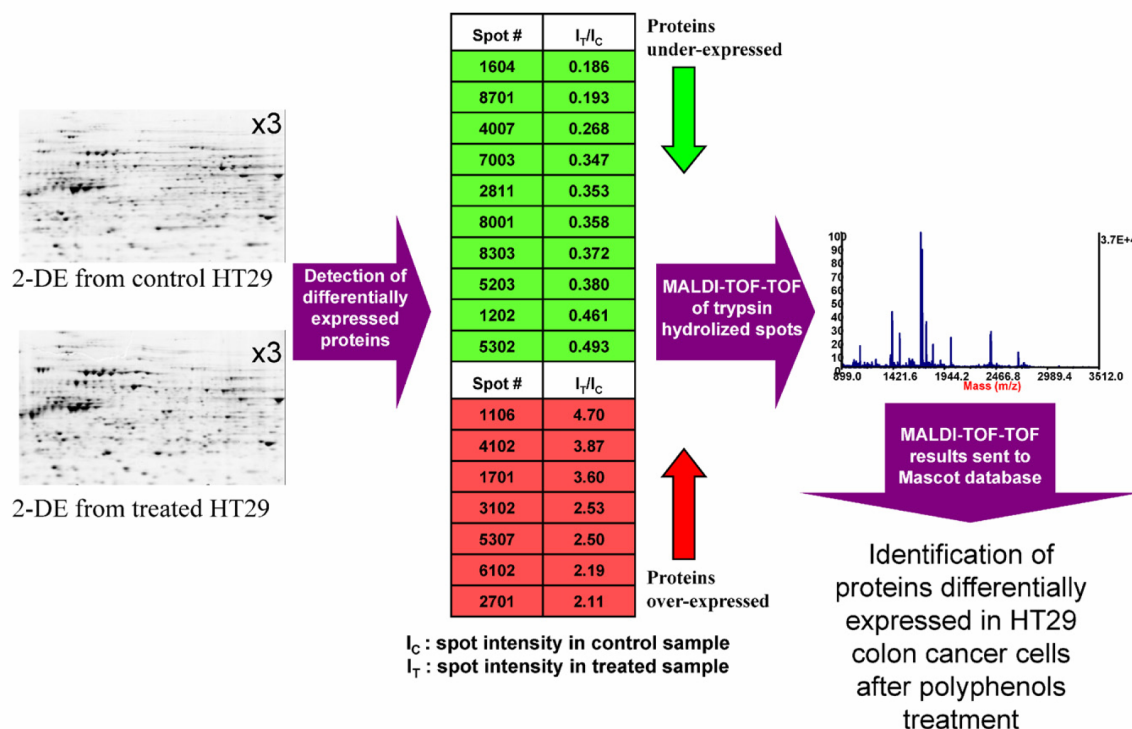
Fig. 5. Scheme of the procedure developed in this work for the Proteomics analysis of HT29 colon cancer cells treated with dietary polyphenols. Minimum three biological replicates were used for each experiment (treated vs. control, i.e., minimum six 2-DE gels were used for each experiment). Differentially expressed proteins were determined by image analysis of the 2-DE gels determining the ratio between the spots intensity in control and treated samples (I_C/I_T, where I_C is the spot intensity in the control sample and I_T in the treated sample). Differentially expressed proteins were hydrolyzed with trypsin and then the peptidic fragments analyzed by MALDI-TOF/TOF. Protein identification was done using the Mascot database.

metabolites given in Table 1 were common, in this case detected by CE-MS and RP/UPLC–MS, providing both techniques the same trend (up-regulation) for the three metabolites. These results show the importance of using several analytical tools simultaneously in Metabolomics to obtain a metabolites coverage as broad as possible.

4. Discussion

4.1. From Transcriptomics results

Transcriptomics serves to put proteomic and metabolomic markers into a larger biological perspective and is suitable for a first "round of discovery" in regulatory networks. Based on transcriptomics results and IPA, a high number of differentially expressed genes in HT29 colon cancer cells were associated with *cellular development, cell death, cellular growth and proliferation* and *cell cycle*, indicating a clear alteration of important biological functions in response to the treatment with rosemary polyphenols. Among the lists of differentially expressed genes, *HMOX1* (GenBank ID: AY460337), *OSGIN1* (GenBank ID: AY258066), *DUSP1* (GenBank ID: X68277.1), *MUC1* (GenBank ID: J05581) and *ARRDC3* (GenBank ID: AB037797) genes were selected for RT-qPCR validation, in order to corroborate the validity of the results obtained from microarrays.

RT-qPCR confirmed that *HMOX1* was significantly up-regulated in HT29 cells by rosemary polyphenols. It is now widely accepted that induction of *HMOX1* expression represents an adaptive response that increases cell resistance to oxidative injury [36,37]. *OSGIN1* is also overexpressed by rosemary compounds being associated with cell growth reduction, increasing apoptosis and

decreasing tumorigenesis, neoplasia or cancer. *MUC1* is down-regulated by rosemary compounds, *MUC1* has been found to be overexpressed in the majority of adenocarcinomas and their precursor lesions, and in various inflammatory diseases [38]. *ARRDC3* is overexpressed by rosemary compounds; it has been seen that expression of *ARRDC3* decreases with tumor grade and metastasis. *ARRDC3* overexpression represses cancer cell proliferation, migration, invasion and growth [39]. *DUSP1* is up-regulated by rosemary polyphenols; it has been observed that *DUSP1* is greatly increased after oxidative stress and also when apoptosis is triggered. On the other hand, overexpression of *DUSP1* activity seems to significantly increase cellular susceptibility to oxidative damage [40].

Transcriptional activation of typical nuclear factor *NRF2* target genes such as *HMOX1*, *TXNRD1*, *GCLM*, *OSGIN1*, *GCLC* and *EPHX1* was observed in HT29 colon cancer cell line by rosemary polyphenols. These enzymes exert efficient cytoprotection against toxicants through a variety of reactions, promoting elimination or inactivation of toxic reactive species (including potential carcinogens) before they cause damage to critical cellular macromolecules [37,41–44].

Transcriptomics data also suggest that rosemary polyphenols may induce endoplasmic reticulum stress-dependent apoptosis in colon cancer cells. However, this pro-apoptotic effect may be counteracted by other mechanism in light to the low degree of apoptosis observed in vitro. IPA results also provided evidences of alteration of G2/M checkpoint control function by modulation of the expression of important genes, including *CDKN1A* and *14-3-3σ* in treated HT29 colon cancer cells. GSEA analysis showed that rosemary polyphenols induces changes in a range of cell cycle processes, accounted by the moderated down-regulation of a number of genes

Table 1
Comparison of the 10 metabolites with the higher fold-change (FC) detected by Metabolomics using RP/UPLC–TOF-MS, CE-TOF-MS and HILIC/UPLC–TOF-MS.
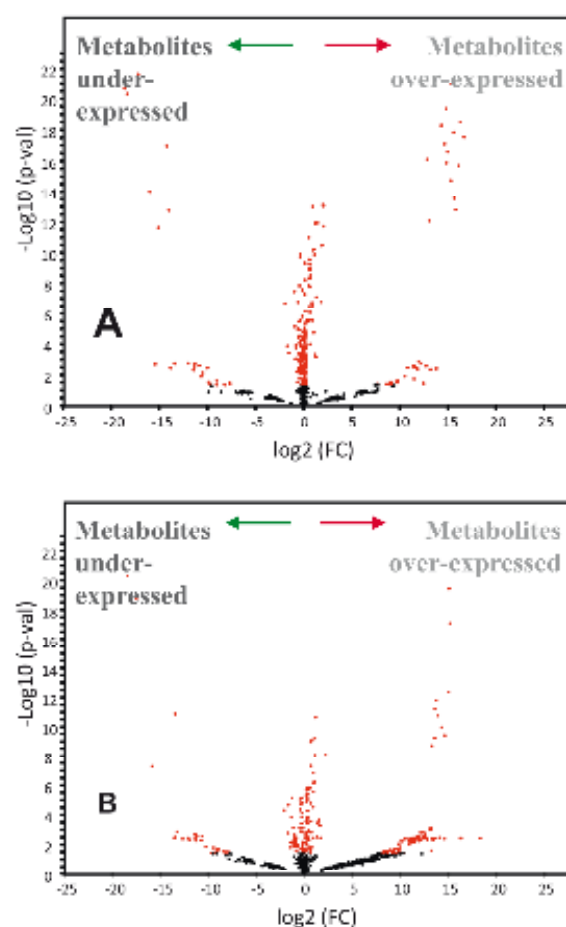
CE-TOF-MS			RP/UPLC–TOF-MS			HILIC/UPLC–TOF-MS		
Measured m/z	Tentative identification	FC	Measured m/z	Tentative identification	FC	Measured m/z	Tentative identification	FC
308.0907	Glutathione, reduced (CID 124886)	>10 ↑	204.1228	Acetylcarnitine [a] (CID 16230)	>10 ↑	136.0618	Adenine (CID 190)	4.6 -
339.1094	Caffeoyl deoxyquinic acid (CID 5281760)	7.7 ↑	192.0809	Tyrosine [a] (CID 6057)	>10 ↓	347.2205	Lys, Ser, Ile [b] (MID 20274) [d]	1.9 -
188.1773	Acetyl-spermidine [a] (CID 496) (CID 123689)	6.3 ↑	253.1185	Tyr, Ala [b] (MID 23276) (MID 21768)	>10 ↑	784.5773	Phosphatidylserine (CID 52920011) [d]	1.6 -
132.1021	Isoleucine [a] (CID 6306)	6.1 ↑	319.1727	Ser, Gly, Asp [c] (MID 18570) [d]	>10 ↑	137.0458	Hypoxanthine (CID 790)	1.4 -
150.0548	Dihydroxyindole (CID 21380)	6.1 ↑	216.1956	N-Nonanoylglycine (CID 10516752)	>10 ↑	884.4724	Phosphatidylethanolamine (CID 9843)	1.3 -
156.0761	Histidine [a] (CID 6274)	6.0 ↑	132.1022	Leucine [a] (CID 6106)	4.2 -	202.1806	Aminoundecanoic acid [a] (CID 17081)	1.3 -
132.1019	Leucine [a] (CID 6106)	5.4 ↑	196.0862	Phenylalanine [a] (CID 6140)	4.0 -	424.1844	Tyr, Gln, Asn [c] (MID 18724) [d]	1.2 -
166.0856	Phenylalanine [a] (CID 6140)	5.4 ↑	335.1377	Glu, Met, Gly [c] (MID 21388) [d]	<0.1 -	341.1082	Ketolactose (CID 97)	1.2 -
182.0802	Tyrosine [a] (CID 6057)	5.3 ↑	555.2984	5-beta-Cyprinolsulfate (CID 440699)	<0.1 -	343.1235	Lactose, Isolactose, Maltose, etc. (CID 84571) [d]	1.1 -
218.1140	Acetyl-citrulline (CID 25245024)	3.7 ↑	220.1177	Pantothenic acid [a] (CID 6613)	<0.1 -	176.0929	Carboxyethyl-gamma-aminobutyric acid (CID 25574)	0.8 -

a Identification confirmed with commercial standards.
b Dipeptide.
c Tripeptide.
d More than two database identities are available. CID: PubChem Compound Identifier; MID: MetLin Metabolite Identifier.

Fig. 6. Volcano plots showing the number of differentially expressed features in treated vs. control colon cancer cells determined by Metabolomics using (A) RP/UPLC–TOF-MS and (B) HILIC/UPLC–TOF-MS.

related with M phase, chromosome segregation and cytokinesis. Further investigations are required to elucidate the nature and the underling mechanisms of the cellular stress induced by the extract. This could be achieved by completing the above transcriptomic results with information from the other two expression levels (i.e., protein and metabolite).

4.2. From Proteomics results

After polyphenols treatment, the up-regulated proteins were identified as cathepsin B (UniProt ID: P07858), cathepsin D (UniProt ID: P07339), peroxiredoxin-4 (UniProt ID: Q13162), phosphoserine phosphatase (UniProt ID: P78330), ribonucleoprotein K (UniProt ID: P61978), serine/threonine-protein phosphatase PP1 (UniProt ID: P62136) and transaldolase 1 (UniProt ID: P37837); while the down regulated proteins were identified as tubulin alpha-1B (UniProt ID: P68363), stathmin A (UniProt ID: P16949), copper-zinc superoxide dismutase (UniProt ID: P00441), chaperonin containing t-complex polypeptide 1 (TCP1) (UniProt ID: P17987), transitional endoplasmic reticulum ATPase (UniProt ID: P55072), fructose-1,6-bisphosphatase (UniProt ID: P09467), glutathione-S-transferase (UniProt ID: P09211) and chain K acetyl-cyclophilin A (UniProt ID: P62937).

Once these proteins were identified, it was observed that their functions were mainly linked to tumorigenesis, cancer proliferation, glycolysis and antioxidant activity. As an example, tubulin alpha-1B is down-regulated by rosemary compounds; this protein has recently been found to be up-regulated in cancer cells more than 2-fold compared to normal cells [45]. Chaperonin containing t-complex polypeptide 1 (TCP-1) is downregulated by rosemary compounds, this chaperonin was found to be over-expressed in colorectal adenocarcinomas [46]. Stathmin is downregulated by rosemary compounds and the overexpression of this protein has been found to promote proliferation, cell adhesion, and migration of human colorectal cancer [47].

However, in other cases the regulation of proteins by the rosemary polyphenols had an unexpected behaviour. Thus, cathepsin B is upregulated by rosemary compounds and this protease has been found to play a critical role in tumorigenesis being upregulated in colorectal cancer [48]. The same applies to cathepsin D, glutathione-S-transferase (an antioxidant enzyme that is down regulated by polyphenols) and ribonucleoprotein K, an essential RNA and DNA binding protein involved in gene expression and signal transduction including DNA transcription, RNA splicing, RNA stability and translation. Although the role of this protein in cancer is relatively understudied, several cellular functions strongly indicate that it is involved in tumorigenesis [49]; in our case, the polyphenols treatment upregulated this protein. Besides, according to literature, the up- or down-regulation of other identified proteins can give rise to opposite results. As an example, peroxiredoxin-4 is also upregulated by rosemary compounds; this protein is an antioxidant enzyme responsible for cell redox homeostasis and has been reported to be overexpressed in colorectal cancer tissue. Increased expression of peroxiredoxin has been linked to resistance to apoptosis under conditions of oxidative stress and this is most likely mediated through the apoptosis signal-regulated kinase-1 signaling pathway. The role of peroxiredoxin and its potential to influence other signaling cascades in colon cancer cell lines needs to be further investigated [50]. Another biological function linked to some detected proteins, namely, transaldolase and fructose-1,6-bisphosphatase was glycolysis. Although it is not clear why transaldolase is upregulated and fructose-1,6-bisphosphatase is dowregulated, it has been shown that elevated glycolysis is linked to higher chemoresistance in colon cancer cells [51]. In summary, Proteomics identified several cellular mechanisms modified by the polyphenols that can explain the antiproliferative activity of these compounds. However, further investigations are required to more clearly elucidate some unexpected behaviors linked to other proteins detected. Complementary information from the other expression levels (i.e., transcript and metabolite) should add more light to understand the main biological mechanisms involved.

4.3. From Metabolomics results

As described in Section 2.7, the total number of tentatively identified metabolites (65) and FC values were uploaded into the IPA software from which 34 metabolites were mapped to biological functions and pathways. 22 metabolites (namely, adenine, adenosine, creatine, reduced glutathione (GSH), oxidized glutathione (GSSG), hypoxanthine, alanine, arginine, histidine, isoleucine, lysine, phenylalanine, tryptophan, tyrosine, N-acetylspermidine, phosphatidylethanolamine, phosphatidylserine, S-adenosyl-L-homocysteine, saccharopine, sarcosine, spermidine and spermine) were highly related metabolites whose change in their intracellular concentration is essential in some cases for the maintenance of the cellular functions. Rosemary polyphenols increased both GSH and GSSG in HT29 cells, although the accumulation rate was higher for GSH (fold change of 17) than for the oxidized form (fold change of 3). Metabolites arginine, creatine,

N-acetyl-L-ornithine, sarcosine, spermidine and spermine were mapped on *urea cycle and metabolism of amino groups* pathway from IPA Knowledge Base (data not shown). Deeper examination of this pathway indicated changes in the levels of intracellular polyamines, namely, spermine and spermidine, upon exposure to rosemary polyphenols. These metabolites regulate important cellular processes including cell proliferation and viability [52], while elevated levels of intracellular polyamines have been associated to certain diseases including colon cancer [53]. Other detected metabolites as S-adenosyl-homocysteine (SAH) [54] or sarcosine [55] have already been suggested as possible biomarkers of different types of cancer and, therefore, their variation can provide interesting information on the activity of the tested polyphenols against colon cancer cells. On the other hand, carnitine, has already been shown to inhibit proliferation and to induce apoptosis in human colon carcinoma cells [56], while high amounts of pantothenic acid in diet has been significantly associated with increased genome instability [57]. Interestingly, in this work the polyphenols treatment induced an increase of carnitine and a decrease of pantothenic acid in colon cancer cells.

4.4. Global Foodomics evaluation: advantages, drawbacks and current challenges

As illustrated in Fig. 1, an ideal Foodomics evaluation on the health benefits of dietary constituents would require integration of the data generated from the three omics levels. Ideally this data integration should follow a systems biology holistic model, which is expected to provide crucial information on biological mechanisms, health effects and relevant biomarkers, in our specific case, on the chemopreventive effect of rosemary polyphenols against colon cancer cells.

However, due to the enormous size and complexity of intracellular biological networks we are still far from a whole-cell computational model that integrates and simulates all the components of a living cell. Besides, computational cell models tend to be partial and focused on an application of interest. Also, due to the multidisciplinarity of the field, these models are based on different kinds of formalisms that hamper the integration of data and models from different sources. Therefore, the whole-cell simulation goal as expected from systems biology is still under development [58]. In addition to this, data integration from the three expression levels (i.e., cross-platform analysis) faces the problem of different timescales of transcript production, protein expression and metabolite accumulation, being the transcriptional response, in general, faster (from seconds to minutes) than proteins (minutes to hours) than metabolites (great time variability). Also, proteins half-life may last from minutes to event months making integrated data not necessarily correlative [2].

An additional difficulty is to consider the different amount of information that is provided by the three omics platforms. In our case, from Transcriptomics, 1308 genes could be identified to be expressed differentially. On the other hand, Metabolomics could identify 65 metabolites changing significantly, while only 17 proteins were found to be differentially expressed and identified by Proteomics. This different amount of information makes even more difficult the possibility to model and handle all this information without a significant bias. This can also explain the difficulty to find in the literature a work in which these three omics technologies are put together to investigate the health benefits of a dietary ingredient (or group of ingredients) at gene, protein and metabolite level.

In this work, an integrative strategy based on the networking and mapping capabilities of IPA software was approached as described in Section 2.8 in order to tackle the cross-platform analysis and biological interpretation. In our case, the integration
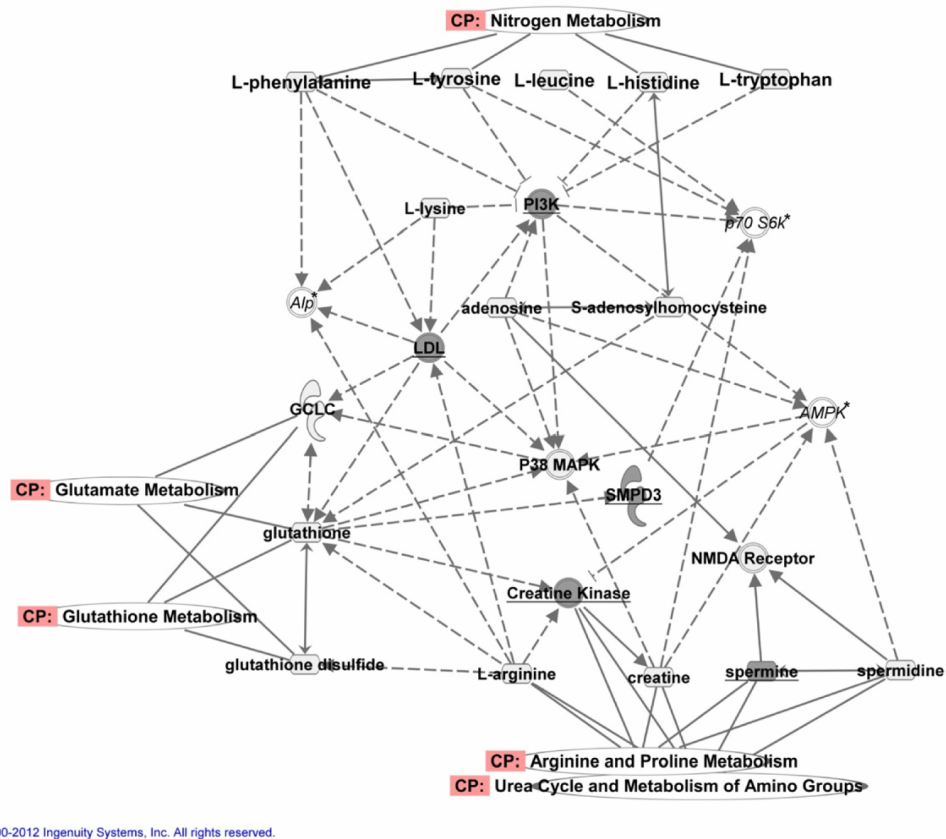
Fig. 7. Top scored metabolite-centric network generated by IPA describing amino acid metabolism, molecular transport and small molecule biochemistry. Microarray results were overlaid in the network highlighting associations between metabolites and transcripts in different canonical pathways (CP). Underlined molecules were down-regulated; not underlined molecules were up-regulated; Alp, p70 S6k and AMPK (marked with an asterisk) were not altered.

of the information from the three levels of expression through IPA did not provide excessive meaningful information. The only direct associations were achieved by overlaying transcriptomic data onto the metabolite-centric network (Fig. 7). Namely, IPA network analysis generated one significant metabolite-centric network (score = 33) including 14 focused metabolites and several associated genes (Fig. 7). The biological functions associated with the network were amino acid metabolism, molecular transport and small molecule biochemistry. The overlay of transcriptomics data on the network illustrated in Fig. 7 reflects that part of the transcriptional response observed in the cells after the treatment may be linked to the altered metabolic pathways. Specifically, glutamate and glutathione metabolism were represented by associations between *GCLC* gene transcript, GSH and GSSG; arginine and proline metabolism, urea cycle and amino groups metabolism were also highlighted by the change in the levels of arginine, creatine, spermidine, spermine, and creatine kinase transcript. On the other hand, using the same approach with proteomics results, no associations were identified between metabolomics and proteomics data. Although the information obtained by this strategy was unexpectedly scarce, the findings motivated us to expand on our observations in these and other related metabolic pathways, such as polyamine metabolism (vide infra).

The main advantage expected from a *Foodomics* evaluation of the health benefits from dietary constituents, is the capability to obtain a global view on their effect at different levels of expression. This can fill the gaps frequently detected when a single omic platform is employed. However, as mentioned above, due to the current limitations in terms of bioinformatic tools and our own limited understanding of the entire cellular processes, the integration of the whole information from the three expression levels cannot be achieved at this stage. Therefore, the following manual approach was proposed: (1) from Transcriptomics, Proteomics and Metabolomics results, the biological processes affected after the treatment with rosemary polyphenols were individually analyzed; (2) overlaps among the three expression levels were detected; and (3) proteins, metabolites and genes linked to these processes were identified within the compounds determined by Transcriptomics, Proteomics and Metabolomics to significantly change. Following this approach, the information obtained from Transcriptomics, Proteomics and Metabolomics was found to overlap over three biological processes. Namely, as illustrated in Fig. 8, polyphenols bring about an induction of cell-cycle arrest, an increase of apoptosis and an improvement of cellular antioxidant activity. The genes, proteins and metabolites identified and implicated in these three processes are shown in Fig. 8 (detected up- and down-regulated compounds are shown in Fig. 8, down-regulated compounds are underlined). Induction of apoptosis is especially relevant in colon cancer, since the renewal of the colon epithelium via apoptosis is the way used by our organism to eliminate deteriorated cells that can mutate to cancerigenous. Thus, significant variations observed for the metabolites arginine, S-adenosylhomocysteine, adenosine and carnitine; together with changes induced in the proteins transitional endoplasmic reticulum ATPase,
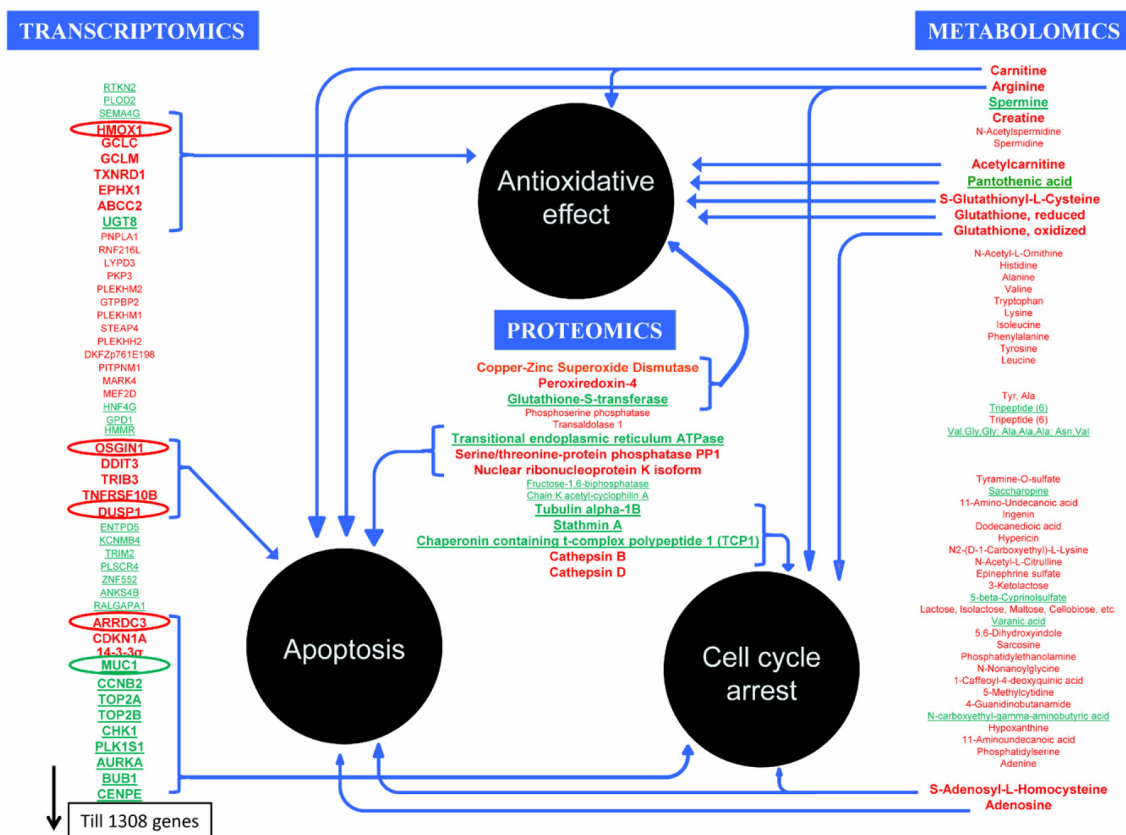
C. Ibáñez et al. / J. Chromatogr. A 1248 (2012) 139–153

Fig. 8. Foodomics identification of the proteins, genes and metabolites involved in three of the principal biological processes altered in HT29 colon cancer cells after the treatment with rosemary polyphenols. Underlined: down-regulated; no underlined: up-regulated.

serine/threonine-protein phosphatase PP1 and nuclear ribonucleo-protein K isoform, and the variations observed in the genes *OSGIN1*, *DDIT3*, *TRIB3*, *TNFRSF10B*, *DUSP1*, can explain the anti-proliferative effect of the rosemary polyphenols against HT29 colon cancer cells due to stress-induced apoptosis. In the same way, cell cycle arrest is also implicated in the anti-proliferative effect from rosemary polyphenols, as can be deduced from changes in S-adenosyl-homocysteine, oxidized glutathione, arginine, tubulin alpha-1B, stathmin A, chaperonin containing t-complex polypeptide 1 (TCP1), cathepsin B, cathepsin D, *ARRDC3*, *CDKN1A*, *14-3-3σ*, *MUC1*, *CCNB2*, *TOP2A*, *TOP2B*, *CHK1*, *PLK1S1*, *AURKA*, *BUB1* and *CENPE* as detected by the three omics technologies used in this work and related to alteration of cell cycle control by polyphenols. Additionally, significant variations observed for metabolites carnitine, acetylcarnitine, reduced glutathione, arginine, pantothenic acid and S-glutathionyl-cysteine; together with changes in proteins copper-zinc superoxide dismutase, peroxiredoxin-4 and, glutathione-S-transferase; in conjunction with the variations detected in genes *HMOX1*, *GCLC*, *GCLM*, *TXNRD1*, *EPHX1*, *ABCC2* and *UGT8*, can explain the chemopreventive effect from rosemary polyphenols via modulation of endogenous antioxidant enzymes and metabolites. Moreover, the relation between each compound and the mentioned biological mechanism could be confirmed through the IPA information and/or literature for all the mentioned cases.

Some of the existing bioinformatic tools allow integrating results from two different expression levels following a more guided process, that is, far from the initial hypothesis-free approach, making possible to study in more detail selected biological processes. To show the possibilities of this guided methodology, we have chosen as examples to study the antioxidant effect and polyamine metabolism mentioned above.

To gain additional information of the antioxidant effect from rosemary polyphenols at the gene expression and protein levels, transcriptomics and proteomics data sets were further analyzed to study with more detail the observed increase of both reduced (GSH) and oxidized (GSSG) glutathione in HT29 cells after the treatment with rosemary polyphenols. As mentioned above, the accumulation rate was higher for GSH than for the oxidized form upon exposure to rosemary polyphenols. Interestingly, transcriptomics provided evidences of the induction of phase II detoxifying and antioxidant genes in response to the treatment with rosemary extract, namely, *HMOX1* (heme oxygenase-1), *OSGIN1* (oxidative stress induced growth inhibitor 1), *TXNRD1* (cytoplasmic thioredoxin reductase-1), *EPHX1* (epoxide hydrolase 1), *GCLM* (glutamate-cysteine ligase, modifier subunit) and *GCLC* (glutamate-cysteine ligase, catalytic subunit). Induction of both, modified and catalytic subunits of glutamate-cysteine genes suggested a potential increase of L-γ-glutamylcysteine biosynthesis. This compound is a precursor used by glutathione synthetase to form GSH. In addition to this, Proteomics results evidenced the expression of the antioxidant proteins copper-zinc superoxide dismutase and peroxiredoxin-4 in treated cells. Thus, the up-regulation of antioxidant proteins and phase II enzymes in treated cells suggest that the phenolic extract may exert cytoprotective effect through a variety of reactions, promoting elimination or inactivation of toxic reactive species, and increasing the intracellular glutathione pool.

PRIMERA PARTE. Capítulo 3

C. Ibáñez et al. / J. Chromatogr. A 1248 (2012) 139–153   151
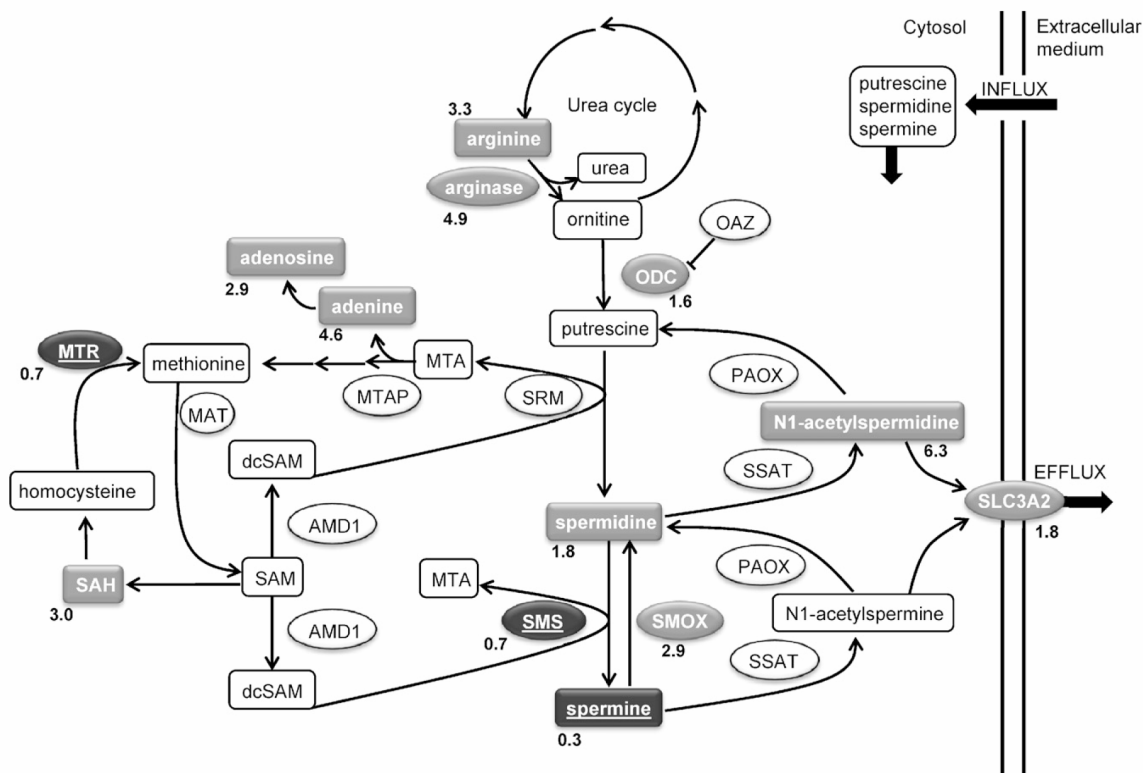
Fig. 9. Polyamine pathway mapping transcriptomics and metabolomics results showing up-regulation of acetylated polyamine efflux, as well as down- and up-regulation of spermine biosynthesis and catabolism, respectively. Oval shape nodes represent the expression of genes and rectangular nodes represent metabolites. Down-regulated components are underlined on a black background; up-regulated are shown on a grey background. Numbers indicate gene/metabolite expression (fold change, natural scale) of the significant altered molecules.

In order to study in more detail the polyamine metabolism, we performed a manually overlaid of metabolomics and transcriptomics data sets on the polyamine pathway, obtaining the results shown in Fig. 9. Thus, spermidine was accumulated (1.8-fold change) more in cells incubated with rosemary polyphenols than in control cells. Coincidently, a significant lower level (0.3-fold change) of spermine was detected in cells incubated with the polyphenols in comparison with the controls. Interestingly, the levels of N-acetylspermidine, the catalytic product of SSAT (spermidine/spermine N1-acetyl-transferase) enzyme, increased greatly in treated cells (6.3-fold change). SSAT is the rate-limiting enzyme of polyamine catabolism and an important regulatory step in maintaining polyamine content. This enzyme adds terminal acetyl groups to spermidine and spermine that subsequently promotes the export of acetylated polyamines. Efflux of acetylated polyamines has been shown to occur in combination with the *SLC3A2* transporter by a polyamine/arginine exchange reaction promoting intracellular polyamine depletion [59,60]. In this work, mRNA levels from *SSAT* gene remained unchanged by the treatment. However, induction of *SLC3A2* gene expression in treated cells provided additional confirmation that rosemary polyphenols may increase the intracellular polyamine depletion by increasing the efflux of their acetylated derivatives. On the other hand, polyamine efflux is often followed by a compensatory increase in polyamine biosynthesis and presumably by an increased flow through the pathway [61]. The spermidine accumulation observed in treated cells supports this idea. Overlay of the transcriptomics results suggested that accumulation of spermidine in treated cells

may be, in part, the result of both, down-regulation of spermine biosynthesis (spermine synthase, *SMS* gene) and increase of spermine conversion to spermidine through induction of spermine oxidase (*SMOX* gene, 2.9-fold change) expression. Interestingly, the intracellular levels of polyamine biosynthesis precursors, ornithine and putrescine remained unaltered after incubation with rosemary extract. However, the induction of gene expression for arginase, a linking node between urea cycle metabolites and polyamine pathway, and to a lesser extent for ornithine decarboxylase (*ODC* gene) may also explain the accumulation of spermidine. The sustained polyamine biosynthesis has been shown to provoke inhibition of cell growth due to the depletion of pathway precursors, such as ornithine, methionine, S-adenosylmethionine (SAM), and acetyl-CoA, and alternatively, due to the overproduction of toxic pathway by-products, such as 5′-methylthioadenosine (MTA), hydrogen peroxide, and aldehydes [62,63]. Regarding polyamine biosynthesis, methionine metabolism plays an important role in providing the methyl group donor SAM that can be decarboxylated by S-adenosyl-methionine decarboxylase 1 to dcSAM, the n-propylamine donor group necessary in spermidine and spermine biosynthesis (see Fig. 9). Metabolomics data highlighted increased levels in some intermediates of the methionine pathway such as adenosine, adenine and S-adenosylhomocysteine (SAH), suggesting also altered profile of the precursors that regulate polyamine biosynthesis [64,65]. Unfortunately, it was not possible to add more evidences on this mechanism based on the proteins level, because none of the 17 proteins identified by Proteomics seem to be involved in this biological process.

5. Conclusions

The results from this work improve our understanding on the chemopreventive effect of rosemary polyphenols against colon cancer cells at molecular level and may be of importance for future prevention and/or treatment of this disease through diet. Besides, our work also provides interesting information about the so-called antioxidant activity of polyphenols. Namely, it is demonstrated that what the investigated polyphenols do is to induce various cellular mechanisms that modify the antioxidant activity inside the cell. These antioxidant properties, together with induction of apoptosis and cell cycle arrest can explain the chemopreventive properties of the mentioned rosemary polyphenols against colon cancer cells.

It is interesting to remark that a number of identified changing genes, proteins and metabolites could not be related to the main biological processes discussed above and, therefore, their effect is not clear at this stage. Besides, no direct correlation could be found between the genes and the proteins whose expression changed more. Last but not least, as shown above, the direct integration following a Systems Biology approach of all the information obtained at the three expression levels is far from being obvious at this moment. The bioinformatic tools capable to handle and integrate all the omics-data generated by different analytical platforms are still distant from what is needed. In summary, this new Foodomics strategy provides an impressive analytical power and it is expected to help to overcome many of the new challenges emerging in Food Science and Nutrition. However, it is observed that together with its remarkable analytical power, new limitations will also come out mainly related to our limited knowledge on the entire cellular mechanisms and the way to handle all the amount of complex multidimensional data and information that can be generated by Foodomics. This point becomes especially critical in those cases in which hypothesis-free approaches want to be applied. A rational and holistic combination of all this information will be crucial in order to extract all the biological meaning from the results provided by this new strategy.

References

[1] M. Muller, S. Kersten, Nat. Rev. Genet. 4 (2003) 315.
[2] M. Kussmann, F. Raymond, M. Affolter, J. Biotechnol. 124 (2006) 758.
[3] Y. Lu, M.V. Boekschoten, S. Wopereis, M. Muller, S. Kersten, Physiol. Gen. 43 (2011) 1307.
[4] A. Cifuentes, J. Chromatogr. A 1216 (2009) 7109.
[5] M. Herrero, V. García-Cañas, C. Simó, A. Cifuentes, Electrophoresis 31 (2010) 205.
[6] M. Herrero, C. Simó, V. García, E. Ibáñez, A. Cifuentes, Mass Spectom. Rev. 31 (2012) 49.
[7] J.R. Araújo, P. Gonçalvesa, F. Martel, Nutr. Res. 31 (2011) 77.
[8] P. Fresco, F. Borges, M.P.M. Marques, C. Diniz, Curr. Pharm. Des. 16 (2010) 114.
[9] S. Ramos, Mol. Nutr. Food Res. 52 (2008) 507.
[10] S.N.T. Ngo, D.B. Williams, R.J. Head, Crit. Rev. Food Sci. Nutr. 51 (2011) 946.
[11] A.B. Granado-Serrano, S. Ramos, Mol. CAB Rev. 35 (2010) 1.
[12] M.B. Gilsenan, Trends Food Sci. Technol. 22 (2011) 536.
[13] G.C.M. Bakker, M.J. van Erk, L. Pellis, S. Wopereis, C.M. Rubingh, N.H.P. Cnubben, T. Kooistra, B. van Ommen, H.F.J. Hendriks, Am. J. Clin. Nutr. 91 (2010) 1044.
[14] S. Ramos, J. Nutr. Biochem. 18 (2007) 427.
[15] E.M.S. McNiven, J.B. German, C.M. Slupsky, J. Nutr. Biochem. 22 (2011) 995.
[16] J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers, D.M. Parkin, Int. J. Cancer 127 (2010) 2893.
[17] A. Trichopoulou, P. Lagiou, H. Kuper, D. Trichopoulos, Cancer Epidem. Biomar. 9 (2000) 869.
[18] V.L. Go, R.R. Butrum, D.A. Wong, J. Nutr. 133 (2003) 3830S.
[19] E.C. Woodhouse, R.F. Chuaqui, L.A. Liotta, Cancer 80 (1997) 529.
[20] H.L. McLeod, J.A. McKay, E.S.R. Collie-Duguid, J. Cassidy, Eur. J. Cancer 36 (2000) 1706.
[21] M. Herrero, M. Plaza, A. Cifuentes, E. Ibáñez, J. Chromatogr. 31 (2010) 205.
[22] R. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Yang, J. Zhang, Genome Biol. 5 (2004) R80.
[23] L. Gautier, L. Cope, B.M. Bolstad, R.A. Irizarry, Bioinformatics 20 (2004) 307.
[24] Y. Benjamini, Y. Hochberg, J. R. Stat. Soc. B 57 (1995) 289.
[25] G.K. Smyth, in: R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (Eds.), Bioinformatics and Computational Biology Solutions using R and Bioconductor, Springer, New York, 2005, p. 397.
[26] M.W. Pfaffl, in: M.T. Dorak (Ed.), Real-time PCR, Taylor & Francis Group, New-castle, 2006, p. 63.
[27] V.K. Mootha, C.M. Lindgren, D.F. Ericsson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M.J. Daly, N. Patterson, J.P. Mesirov, T.R. Golub, P. Tamayo, B. Spiegelman, E.S. Lander, J.N. Hirschhorn, D. Altshuler, L.C. Groop, Nat. Genet. 34 (2003) 267.
[28] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545.
[29] A. Shevchenko, M. Wilm, O. Vorm, M. Mann, Anal. Chem. 68 (1996) 850.
[30] C.A. Smith, E.J. Want, G.C. Tong, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.
[31] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, Nucleic Acids Res. 37 (2009) D603.
[32] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, Ther. Drug Monit. 19 (2005) 258.
[33] C.D. Davis, J. Milner, Mutat. Res. 551 (2004) 51.
[34] V. García-Cañas, C. Simó, C. León, A. Cifuentes, J. Pharm. Biomed. 51 (2010) 290.
[35] V. García-Cañas, M. Mondello, A. Cifuentes, Electrophoresis 31 (2010) 2249.
[36] C.K. Andreadi, L.M. Howells, P.A. Atherfold, M.M. Manson, Mol. Pharm. 69 (2006) 1033.
[37] D. Martin, A.I. Rojo, M. Salinas, R. Diaz, G. Gallardo, J. Alam, C.M. Ruiz de Galarreta, A. Cuadrado, J. Biol. Chem. 279 (2004) 8919.
[38] P.L. Beatty, S. Narayanan, J. Gariépy, S. Ranganathan, O.J. Finn, Cancer Prev. Res. 3 (2010) 438.
[39] K.M. Draheim, H.B. Chen, Q. Tao, N. Moore, M. Roche, S. Lyle, Oncogene 29 (2010) 5032.
[40] Y.X. Liu, J.L. Wang, J.F. Guo, J.J. Wu, H.B. Lieberman, Y.X. Yin, Mol. Cancer Res. 6 (2008) 624.
[41] O.I. Aruoma, B. Halliwell, R. Aeschback, J. Lolingers, Xenobiotica 22 (1992) 257.
[42] W.S. Jeong, M. Jun, A.N.T. Kong, Antiox. Red. Sign. 8 (2006) 99.
[43] O. Rau, M. Wurglics, A. Paulke, J. Zitzkowski, N. Meindl, A. Bock, T. Dingermann, M. Abdel-Tawab, M. Schubert-Zsilavecz, Planta Med. 72 (2006) 881.
[44] J.W. Kaspar, S.K. Niture, A.K. Jaiswal, Free Radical Bio. Med. 47 (2009) 1304.
[45] J. Hye-Cheol, K. Gwang-Il, C. Sang-Ho, L. Kwang-Hyung, K. Jung-Jae, Y. Jeong-Hee, C. Kwang-Hoe, J. Proteome. Res. 10 (2011) 269.
[46] C. Coghlin, B. Carpenter, S.R. Dundas, L.C. Lawrie, C. Telfer, G.I. Murray, J. Pathol. 210 (2006) 351.
[47] P. Zheng, Y.X. Liu, L. Chen, X.H. Liu, Z.Q. Xiao, L. Zhao, G.Q. Li, J. Zhou, Y.Q. Ding, J.M. Li, J. Proteome. Res. 9 (2010) 4897.
[48] A.T. Chan, Y. Baba, K. Shima, K. Nosho, D.C. Chung, K.E. Hung, U. Mahmood, K. Madden, K. Poss, A. Ranieri, D. Shue, R. Kucherlapati, C.S. Fuchs, S. Ogino, Cancer Epidem. Biomar. 19 (2010) 2777.
[49] R. Zhou, R. Shanas, M.A. Nelson, A. Bhattacharyya, J. Shi, Int. J. Cancer 126 (2010) 395.
[50] K. Fung, G.V. Brierley, S.T. Henderson, P. Hoffmann, S.R. McColl, T. Lockett, R. Head, L.J. Cosgrove, J. Proteome. Res. 10 (2011) 1860.
[51] Y. Zhou, F. Tozzi, J. Chen, F. Fan, L. Xia, J. Wang, G. Gao, A. Zhang, X. Xia, H. Brasher, W. Widger, L.M. Ellis, Z. Weihua, Cancer Res. 72 (2012) 304.
[52] E.W. Gerner, F.L. Meyskens, Nat. Rev. 4 (2004) 781.
[53] E.A. Paz, J. García-Huidobro, N.A. Ignatenko, Adv. Clin. Chem. 54 (2011) 45.
[54] J.F. Leal, I. Ferrer, C. Blanco-Aparicio, J. Hernández-Losa, S. Ramón y Cajal, A. Carnero, M.E. Lleonart, Carcinogenesis 29 (2008) 2089.
[55] X. Chen, R. Overcash, T. Green, D. Hoffman, A.S. Asch, M.J. Ruiz-Echevarría, J. Biol. Chem. 286 (2011) 16091.
[56] M.J. Roy, S. Dionne, G. Marx, I. Qureshi, D. Sarma, E. Levy, E.G. Seidman, Nutrition 25 (2009) 1193.
[57] M. Fenech, P. Baghurst, W. Luderer, J. Turner, S. Record, M. Ceppi, S. Bonassi, Carcinogenesis 26 (2005) 991.
[58] A.A. De Graaf, A.P. Freidig, B. De Roos, N. Jamshidi, M. Heinemann, J.A.C. Rullmann, K.D. Hall, M. Adiels, B. Van Ommen, PLoS Comput. Biol. 5 (2009) e1000554.
[59] T. Uemura, H.F. Yerushalmi, G. Tsprailis, D.E. Stringer, K.E. Pastorian, L. Hawell, C.V. Byus, E.W. Gerner, J. Biol. Chem. 283 (2008) 26428.
[60] T. Uemura, D.E. Stringer, K.A. Blohm-Mangone, E.W. Gerner, Am. J. Physiol. Gastrointest. Liver Physiol. 299 (2010) G517.

[61] D.L. Kramer, P. Diegelman, J. Jell, S. Vujcic, S. Merali, C.W. Porter, J. Biol. Chem. 283 (2008) 4241.

[62] H.C. Ha, P.M. Woster, J.D. Yager, R.A. Casero, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 11557.

[63] S. Vujcic, M. Halmekyto, P. Diegelman, G. Gan, D.L. Kramer, J. Janne, C.W. Porter, J. Biol. Chem. 275 (2000) 38319.

[64] A.L. Subhi, P. Diegelman, C.W. Porter, B. Tang, Z.J. Lu, G.D. Markham, W.D. Druger, J. Biol. Chem. 278 (2003) 49868.

[65] A.P. Stevens, K. Dettmer, G. Kirovski, K. Samejima, C. Hellerbrand, A.D. Bosser-hoff, P.J. Oefner, J. Chromatogr. A 1217 (2010) 3282.

3.5.   CONCLUSIONES

En el trabajo descrito en el **Capítulo 1** se presenta un método analítico mediante CE-TOF MS para el análisis de metabolitos procedentes de células humanas modelo de cáncer de colon HT-29. El método desarrollado permitió la obtención de un perfil metabólico en menos de 20 minutos de análisis, y con una buena reproducibilidad. Los valores de RSD fueron inferiores al 12% y al 15% en los estudios de reproducibilidad llevados a cabo en el mismo día (n= 5 inyecciones) y en tres días diferentes (n=15 inyecciones), respectivamente.

Con el fin de estudiar el efecto que tiene el tratamiento de la muestra previo al análisis mediante CE-MS, se estudiaron cuatro metodologías de purificación diferentes de la fracción metabólica de células HT-29.

Al comparar los resultados obtenidos del análisis mediante CE-MS de los dos extractos metabólicos obtenidos mediante SPE (sorbente ABN *vs.* C18) se observaron ligeras diferencias en el número y naturaleza de los metabolitos. Así, empleando la columna con el sorbente ABN se detectaron 80 compuestos diferentes, mientras que usando la extracción con un sorbente del tipo C18 se detectaron 71 compuestos. A pesar de las diferentes propiedades físico-químicas de ambos sorbentes, 62 compuestos detectados fueron sustancias comunes a ambos extractos. Por otro lado, se puede destacar un ligero mayor poder de extracción del sorbente ABN.

Cuando se analizaron los extractos purificados mediante dos procesos diferentes de desproteinización, se obtuvo un número total de metabolitos igual a 83 empleando la precipitación de proteínas con MeOH y de 74 empleando ultrafiltración con membranas de 3 kDa. Al comparar los resultados, de todos los compuestos detectados en ambos extractos mediante CE-MS, 72 de ellos fueron comunes a los dos extractos, por lo que en este caso se puede destacar la similitud de ambos métodos de purificación.

Las diferencias más notables se obtuvieron al comparar los perfiles metabólicos de los extractos más ricos, es decir, el obtenido empleando SPE con sorbente del tipo ABN (80 metabolitos) y la precipitación de proteínas con metanol (83 metabolitos).

Solamente se encontraron 48 compuestos comunes, lo cual demuestra la importancia del tratamiento previo de la muestra, el cual siempre conlleva una cierta selectividad hacia un grupo determinado de compuestos.

En el **Capítulo 2** se ha estudiado el efecto de un extracto rico en polifenoles procedente de romero con potencial actividad biológica en células de cáncer de colon HT-29.

El extracto de romero obtenido mediante la tecnología de fluidos supercríticos y que presentaba carnosol y ácido carnósico como los fitoquímicos mayoritarios, mostró un efecto antiproliferativo al tratar las células HT-29 durante 72 h y una concentración de 10 μM de polifenoles totales.

Se ha estudiado a nivel metabolómico el efecto antiproliferativo observado al tratar en las condiciones mencionadas anteriormente las células HT-29 con el extracto de romero rico en polifenoles. Con el fin de obtener la mayor información metabolómica posible de las células HT-29 se ha combinado el uso de diferentes plataformas analíticas, concretamente, RP/UPLC-TOF MS, HILIC/UPLC-TOF MS y CE-TOF MS.

La combinación de los resultados obtenidos mediante estas técnicas analíticas complementarias permitió aumentar la detección del número de metabolitos cuya expresión era significativamente diferente en células HT-29 tras el tratamiento con el extracto de romero. El tratamiento de las células con este extracto rico en polifenoles reveló un aumento en el ratio entre el glutatión reducido y oxidado (GSH/GSSG) y una alteración en el contenido de poliaminas intracelulares de reconocida importancia en procesos involucrados en la proliferación de células de cáncer.

En el **Capítulo 3** se presenta una estrategia Foodómica como herramienta para la evaluación global del efecto potencialmente beneficioso del extracto de romero rico en polifenoles empleado en el **Capítulo 2,** en células de cáncer de colon HT-29. La combinación de distintas tecnologías ómicas (Transcriptómica, Proteómica y Metabolómica) nos ha permitido profundizar en los mecanismos a nivel molecular modificados tras el tratamiento con el extracto.

Se observó una alteración significativa en la expresión de 1308 genes y 17 proteínas como consecuencia del tratamiento con el extracto polifenólico.

A través del estudio metabolómico comparativo se observaron diferencias estadísticamente significativas (p<0.05) en 210 (RP/UPLC-MS), 214 (HILIC/UPLC-MS) y 212 (CE-MS) metabolitos, de los que 65 pudieron ser identificados tentativamente.

Empleando el software "Ingenuity Pathway Analysis" (IPA) se integraron los resultados de los estudios metabolómicos y transcriptómicos, observándose una alteración en funciones biológicas y rutas metabólicas fuertemente ligadas al mantenimiento de las actividades celulares. Concretamente IPA reveló alteraciones en el metabolismo del nitrógeno, del glutamato, del glutatión, de la arginina y la prolina, y también del ciclo de la urea y metabolismo de grupos amino. La integración directa de los resultados de los tres niveles de expresión mediante el software IPA no proporcionó información relevante. Entre otras razones esto se debió a: i) el reducido número de proteínas detectadas como significativamente diferentes tras el tratamiento con romero; ii) la ausencia de identificadores para muchos metabolitos en el software IPA; iii) el desconocimiento aún existente de muchos de los procesos celulares; iv) la diferente cinética de los procesos que tiene lugar a nivel transcriptómico, proteómico y metabolómico; etc. A pesar de estas limitaciones, la aproximación seguida integrando las rutas alteradas entre los niveles de expresión tomados de dos a dos y el posterior estudio de sus conexiones reveló que la actividad antiproliferativa del extracto frente a células de cáncer de colon se puede relacionar con un aumento de la apoptosis, un aumento de la actividad antioxidante dentro de la célula (mediante la generación de proteínas de fase II) y un aumento del arresto del ciclo celular.

Esta investigación amplía el conocimiento del efecto quimiopreventivo y actividad antioxidante de los polifenoles del extracto de romero en células de cáncer de colon HT-29. Además en este trabajo se ponen de manifiesto las grandes limitaciones y retos en Foodómica, relacionados con el reducido conocimiento actual en cuanto a los mecanismos celulares, y la gran dificultad en el procesamiento e integración de datos procedentes de los tres niveles de expresión (Transcriptómica, Proteómica y Metabolómica).

4.   SEGUNDA PARTE

BÚSQUEDA DE NUEVOS BIOMARCADORES RELACIONADOS CON LA PROGRESIÓN DE LA ENFERMEDAD DE ALZHEIMER

4.1. INTRODUCCIÓN

Existe un gran número de enfermedades, cuya incidencia aumenta de forma pronunciada con la edad, entre las que se encuentra la enfermedad de Alzheimer (AD), que supone el tipo de demencia más prevalente entre la población de edad avanzada (Blennow y col., 2006). La AD es una patología multifactorial que afecta al sistema nervioso central, en el cual se produce una lenta destrucción y atrofia de la corteza cerebral. Las regiones cerebrales que se asocian con funciones mentales superiores, particularmente el neocórtex y el hipocampo, son las más afectadas (Francis y col., 1999). En estas regiones se produce una degeneración neurofibrilar dando lugar a los denominados "ovillos neurofibrilares" (NFT), la formación de placas amiloides y la aparición de infiltrados amiloides en la microvasculatura cerebral. Como consecuencia se produce la muerte neuronal lo que origina una alteración progresiva de las funciones intelectuales a través de un grave deterioro de las funciones cognitiva, conductual y funcional. Estas alteraciones en el sistema nervioso central presentan una sintomatología que comienza con trastornos leves de memoria y deriva en una demencia grave de tipo cortical.

La demencia asociada a Alzheimer conlleva graves repercusiones, no sólo para el enfermo y sus familiares, si no para el conjunto de la sociedad con un gasto sanitario muy elevado, calculado en 2005 en 604 billones de dólares (464 billones de euros aproximadamente) (Wimo y Prince, 2010). Dadas las repercusiones sociales y la creciente esperanza de vida estimada para las próximas generaciones, la investigación sobre los mecanismos mediante los que se desarrolla esta patología tiene en la actualidad una gran importancia. Aunque el primer caso de AD se descrbió hace más de 100 años (en el año 1901) y a pesar de las numerosas investigaciones realizadas hasta la fecha, actualmente no se conocen las causas (Swerdlow, 2012) que preceden a la neurodegeneración patológica descrita por Alois Alzheimer (Alzheimer, 1907). Parece cada vez más evidente que se trata del resultado de la combinación de factores de riesgo genéticos, ambientales y derivados del estilo de vida (Povova y col., 2012). Hasta la fecha, se han postulado diversas hipótesis acerca de la etiología de la enfermedad asociadas con el déficit de acetilcolina, con la acumulación de proteínas o con trastornos metabólicos de distinta naturaleza. Las teorías que han presentado una mayor relevancia, se describen a continuación.

La "**hipótesis colinérgica**" o "**déficit colinérgico cortical**" fue la primera teoría neuroquímica propuesta para explicar el origen de la enfermedad (Bartus y col., 1982; Bartus,

119

2000). Se basa en el descubrimiento de la alteración en los niveles de las enzimas responsables de la síntesis y degradación de la acetilcolina (ACh) y una notable disminución en los niveles de captación neuronal de la ACh (Rylett y col., 1983) generando un déficit de dicho neurotranmisor en el cerebro de pacientes con AD. Según esta hipótesis el origen de la enfermedad se produce como consecuencia de la deficiencia presináptica en los niveles de ACh, y en menor medida de noradrenalina y serotonina en la zona del hipocampo, que se generaliza posteriormente a todo el cerebro. Se han realizado múltiples estudios experimentales desde la aparición de esta hipótesis como explicación del origen de AD, así como estudios farmacológicos con inhibidores de la acetilcolinesterasa. Como resultado de dichos experimentos se ha asociado la proporción de un posible efecto neuroprotector con la administración de fármacos procolinérgicos (Manzano-Palomo y col., 2006). Sin embargo, las críticas más extendidas en contra de esta teoría apuntan que esta hipótesis no puede explicar por sí sola la patogenia asociada a AD ya que los fármacos procolinérgicos no son un tratamiento curativo eficaz (Dumas y Newhouse, 2011). Además, se han detectado otros sistemas de neurotransmisión alterados, y por otro lado, la hipótesis colinérgica no explica las bases moleculares y los factores genéticos por los que se desencadena la enfermedad (Wenk, 2003).

La gran mayoría de marcadores bioquímicos relacionados con la AD han sido fruto de estudios basados en la "**hipótesis amiloide**" que asocia la enfermedad a una acumulación anormal de moléculas proteicas β-amiloide (Aβ) en el cerebro (Tam y Pasternak, 2012; Mullane y Williams, 2013). Esta hipótesis afirma que existe una disfunción en el procesado de la proteína precursora amiloidea (APP) que se encuentra en la superficie de las neuronas. De forma mayoritaria, la proteína APP sufre el procesamiento no amiloidogénico (Figura 4.1). Inicialmente la APP es digerida por enzimas secretasas tipo alfa (α) generando un péptido extracelular llamado sAPPα y una cadena peptídica intermembrana. Sobre esta cadena actúa la secretasa gamma (γ) generando un péptido extracelular soluble (p3) y uno intracelular llamado AICD (siglas en inglés del dominio intracelular de la APP). El proceso amiloidogénico (menos común aunque también fisiológico) se produce cuando en vez de actuar inicialmente la isoforma α de la secretasa, actúa la β que genera un péptido extracelular llamado sAPPβ y uno intermembrana sobre el que actúa la secretasa γ (Figura 4.1). El resultado es el péptido Aβ (con menor solubilidad que el p3 en el espacio extracelular) y una molécula de AICD. Los péptidos β-amiloides generados en el espacio extracelular tienen la capacidad de formar agregados de oligómeros, de 6 a 8 unidades de péptidos Aβ, con un total de 40 aminoácidos (4 kDa de peso

molecular). Según esta teoría, la acumulación y depósito de dichos péptidos amiloides es tóxico para las neuronas e induce su muerte. Las placas amiloides se producen por la acumulación de dichos péptidos de tipo beta amiloide (Aβ) insolubles y densos en el espacio interneuronal. El principal problema que presenta esta teoría, es que la presencia de placas amiloides no es un síntoma patognómico de AD por lo que no puede ser la única causa por la que aparece la enfermedad. Además todavía se desconoce si la acumulación de Aβ es debida a una síntesis excesiva (por un incremento de la acción de la β-secretasa) o a una alteración en los mecanismos de depuración de los péptidos Aβ.
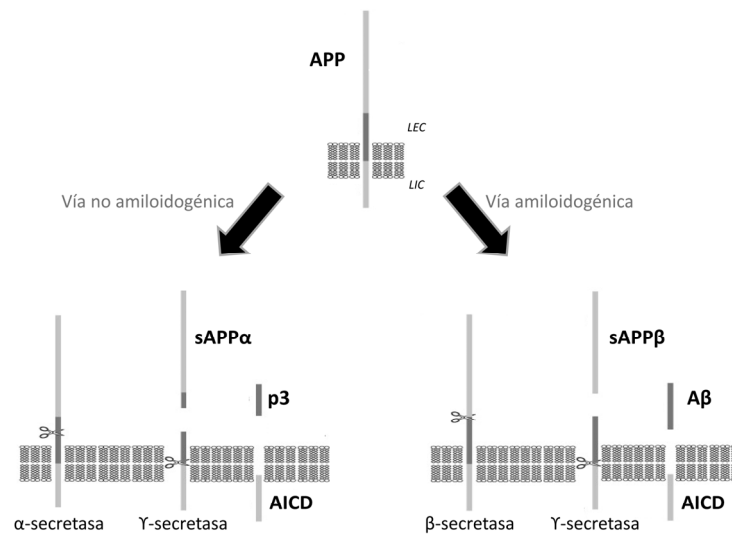
Figura 4.1. Proceso proteolítico por el que se genera el precursor de la proteína amiloide (APP).

Otra de las teorías más importantes en la investigación de la patología de Alzheimer es la "**hipótesis tau**" (Hardy y Allsop, 1991). Se basa en la existencia de "ovillos neurofibrilares" (NFT) compuestos por filamentos dobles dispuestos en espiral (filamentos dobles apareados) formados por proteínas del citoesqueleto neuronal, de entre las que destaca la proteína tau (τ). En condiciones fisiológicas, ciertos aminoácidos de esta proteína citoplasmática se encuentran fosforilados. La hiperfosforilación de la proteína τ hace que se disocie de los microtúbulos citoplasmáticos produciendo por un lado un agregado de proteínas τ hiperfosforiladas insolubles que componen los ovillos neurofibrilares, y por otro lado, la ruptura de la estructura microtubular (Figura 4.2). Este hecho provoca en última instancia la destrucción del citoesqueleto celular y la consiguiente muerte neuronal. Cuando esto sucede, se observan "ovillos fantasmas" en el espacio interneuronal. A diferencia de las placas amiloides, los ovillos neurofibrilares se consideran un síntoma casi patognómico de la enfermedad de

Alzheimer, ya que sólo se han observado en pacientes con otras enfermedades cerebrales con mucha menor incidencia como son la parálisis supranuclear, la demencia frontotemporal y en algunos casos, con la enfermedad de Creutzfeldt-Jakob (Von Bernhardi-Montgomery, 2005).
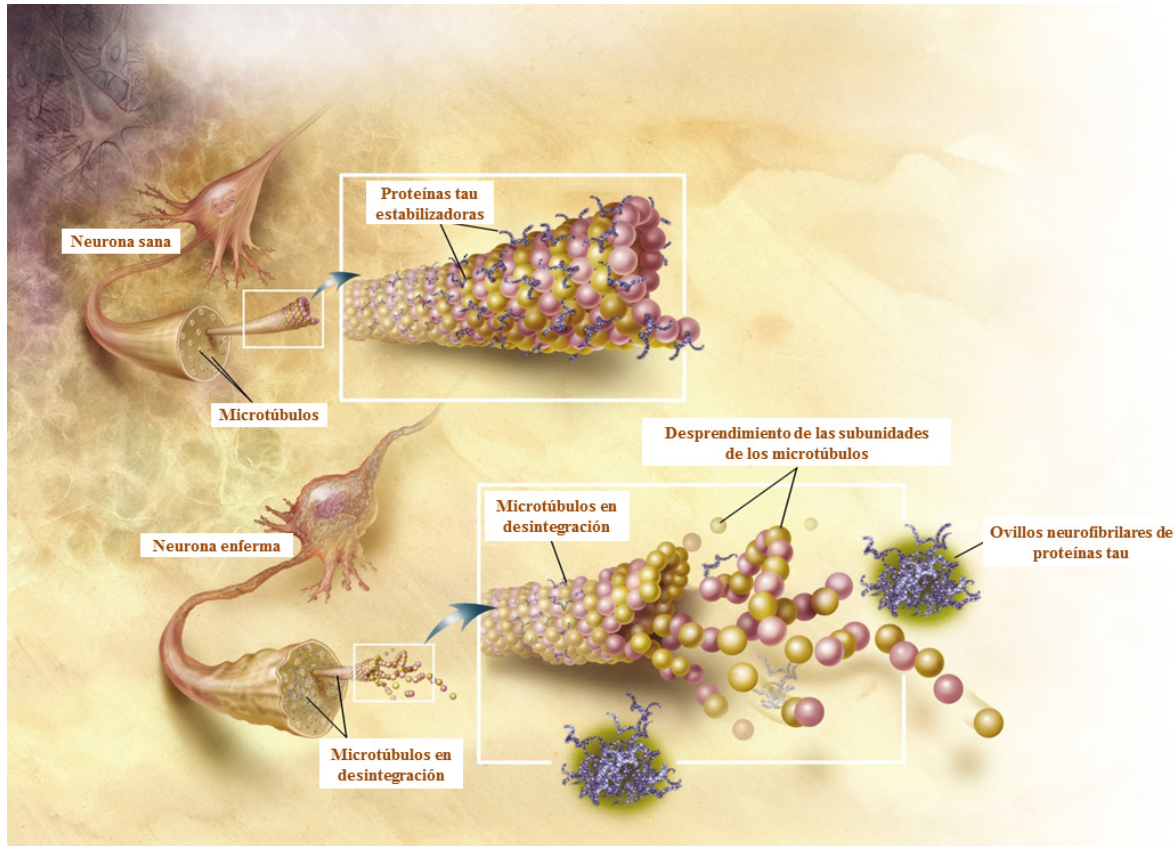
Figura 4.2. Esquema de la formación de los ovillos neurofibrilares y desintegración de los microtúbulos del citoesqueleto celular.

Nota: Imagen modificada a partir de la imagen original procedente de la siguiente página web:

http://www.javeriana.edu.co/Facultades/Ciencias/neurobioquimica/libros/neurobioquimica/cronica13_2.jpg

Lo cierto es que hoy en día, no es posible dilucidar si las moléculas implicadas en las hipótesis anteriores se pueden considerar iniciadores de la patología o son más bien consecuencia de otras alteraciones previas (Jellinger y Bancher, 1998). En cualquier caso la aparición de ovillos neurofibrilares y/o de placas amiloides definen las evidencias más características de la enfermedad y su presencia es clave en el diagnóstico del enfermo de AD. De hecho, la única herramienta que garantiza un diagnóstico definitivo es el examen microscópico cerebral *post mortem* (ver Figura 4.3), en el que se pueden observar las placas amiloides en el espacio interneuronal y los ovillos neurofibrilares en el interior de las neuronas (Tiraboschi y col., 2004). Dado el desconocimiento de la causalidad etiológica y mecanismos fisiopatológicos que gobiernan esta enfermedad parece evidente la ausencia de herramientas de

diagnóstico eficaces en etapas tempranas de la enfermedad (demencia leve o moderada e incluso fases presintomáticas).
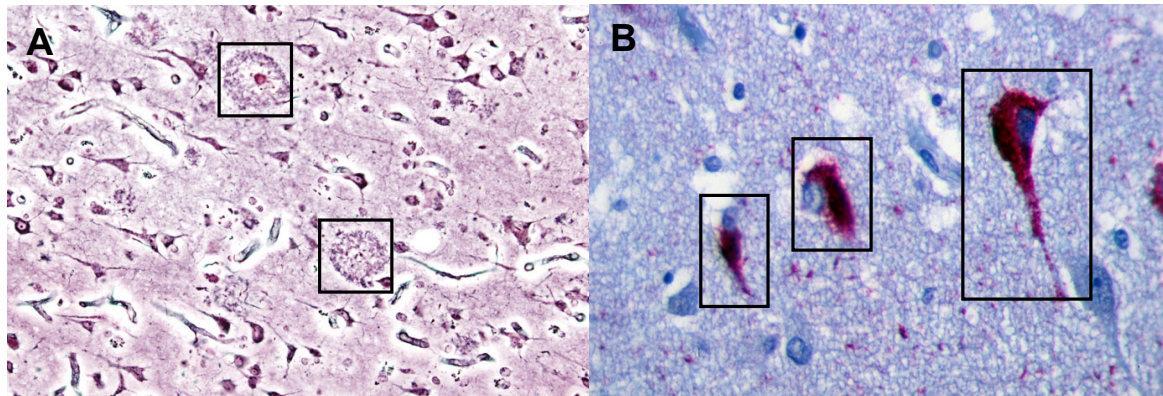
Figura 4.3. Imagen histológica de córtex cerebral de un paciente con la enfermedad de Alzheimer donde se observan (A) placas seniles y (B) ovillos neurofibrilares, ambos enmarcados en negro en la figura.

Nota: Imágenes modificadas a partir de las imágenes originales procedentes de las siguientes direcciones web:

(A) http://25.media.tumblr.com/tumblr_lrohzwCufi1qzcf71o1_500.jpg.

(B) http://upload.wikimedia.org/wikipedia/commons/a/a6/Neurofibrillary_tangles_in_the_Hippocampus_of_an_old_person_with_Alzheimer-related_pathology%2C_immunohistochemistry_for_tau_protein.JPG.

El diagnóstico rutinario de AD comienza con un estudio de la historia clínica del paciente, en búsqueda de antecedentes familiares, y en un examen neuropsicológico para determinar el nivel cognitivo. Si el paciente muestra cierto deterioro cognitivo se le somete a pruebas de imagen cerebral entre las que predominan la resonancia magnética nuclear, la tomografía por emisión de positrones (PET) y la tomografía computarizada por emisión de fotones individuales (SPECT) (Reiman y Jagust, 2012; Román y Pascual, 2013). Los hallazgos típicos son los de atrofia cortical, fundamentalmente a nivel temporal y, más concretamente, en los hipocampos. Aunque existen evidencias de la existencia de una fase presintomática de AD (Sperling y col., 2011), hasta la fecha el deterioro cognitivo leve o demencia incipiente (MCI) es el estadío más temprano conocido de AD. Por este motivo la gran mayoría de las investigaciones en busca de biomarcadores tempranos de la enfermedad se centran en el estudio de la progresión desde MCI a AD, teniendo siempre en cuenta que no todos los casos con MCI progresarán a AD (Chong y Sahadevan, 2005). Los biomarcadores de origen protéico son los más conocidos y utilizados para el diagnóstico de AD, ya que han sido ampliamente estudiados (Shi y col., 2009; Lista y col., 2013). Mediante la determinación de los niveles en las proteínas $A\beta$ y τ, en combinación con las técnicas de imagen, se llega a diagnosticar el 90% de pacientes con AD avanzado (Cedazo-Minguez y Winblad, 2010), con el inconveniente de suponer un elevado coste por cada paciente (6000-10000 euros aproximadamente).

El desarrollo de metodologías novedosas en el descubrimiento de nuevos biomarcadores fiables y eficaces en el diagnóstico precoz de AD se impone como una necesidad cada vez más urgente. La prioridad actual en la investigación de AD se centra en la identificación de biomarcadores que permitan un diagnóstico capaz de reconocer los primeros síntomas de la enfermedad (o incluso antes de que se produzcan síntomas cognitivos), lo que contribuirá a revelar los mecanismos por los que se inicia y progresa la neurodegeneración, desconocidos hasta ahora (Bazenet y Lovestone, 2012), permitiendo de este modo el tratamiento precoz de la enfermedad. Ante el gran desconocimiento acerca del Alzheimer el empleo de las tecnologías ómicas en el estudio de esta enfermedad puede dar lugar a nuevas hipótesis no contempladas anteriormente. Además del empleo de estrategias analíticas no dirigidas como la Transcriptómica (Sutherland y col., 2011) y la Proteómica (Shi y col., 2009; Lista y col., 2013), se observa un creciente interés en la aplicación de estrategias metabolómicas no dirigidas en la investigación de enfermedades neurológicas en general y de AD en particular (Hassan-Smith y col., 2012). Así, la Metabolómica emerge como un enfoque alternativo que ofrece nuevas posibilidades, de modo que, a pesar de encontrarse en su infancia en la investigación de enfermedades neurodegenerativas (Hassan-Smith y col., 2012), la Metabolómica se ha aplicado con éxito en algunos estudios relacionados con la búsqueda de nuevos biomarcadores de AD, tal y como se recoge en la Tabla 4.1. El tejido cerebral (*post mortem*) y el CSF han sido las muestras en las que se han realizado un mayor número de estudios metabolómicos, principalmente por formar parte del sistema nervioso central (CNS). El tejido cerebral ha sido la muestra más usada tradicionalmente en la investigación de AD, gracias a la cual se han desarrollado importantes fármacos e hipótesis etiológicas (Beach, 2013). Por otro lado el CSF es el único fluido biológico cuya composición se ve afectada directamente por el CNS, y por tanto en el que se esperan mayores cambios a nivel bioquímico asociados a AD. Desde el punto de vista clínico, hay una clara tendencia hacia el análisis de otras muestras menos invasivas (por ejemplo, sangre u orina). En este sentido muchas de las investigaciones se están dirigiendo hacia la búsqueda de biomarcadores en este tipo de muestras biológicas (Aftab y Waraich, 2012).

Tabla 4.1. Artículos en los que se estudia la enfermedad de Alzheimer mediante estrategias metabolómicas publicados hasta la fecha (Mayo de 2013) según la base de datos ISI Web of Knowledge.

Muestra	Tipo de análisis	Técnica analítica	Referencia
CSF Sujetos sanos con perfil de riesgo (Aβ y τ) de padecer AD (n=10) y sin perfil de riesgo (n=34).	Huella metabólica	NMR	Jukarainen y col., 2008
Suero MCI (n=19) y controles (n=26).	Perfil de lípidos	NMR	Tukiainen y col., 2008
Plasma Pacientes con AD (n=16), MCI (12) y controles (n=10).	Perfil de ácidos biliares	UPLC-MS/MS	Greenberg y col., 2009
Tejido post-mortem de ratones 8 regiones cerebrales de ratones jóvenes TgCRND8 (n=5) y controles (n=4) y ratones en edad avanzada TgCRND8 (n=3) y controles (n=5).	Huella metabólica	NMR	Salek y col., 2010
Tejido post-mortem Pacientes con AD (n=30) y controles (n=14). **Tejido post-mortem de ratones** Ratones APPswedish (n=5) y controles (n=8).	Análisis diana de tres fosfolípidos	FIA[a]-MS/MS	Grimm y col., 2011
Plasma Pacientes con AD (n=26) y controles (n=26).	Perfil de lípidos	QqQ-MS/MS	Han y col., 2011
CSF post-mortem Pacientes con AD (n=15) y controles (n=15).	Huella metabólica	HPLC-ECA[b]	Kaddurah-Daouk y col., 2011
Suero Pacientes con AD (n=37), MCI con progresión a AD (n=52), MCI sin progresión a AD (n=91) y controles (n=46).	Perfil de lípidos Huella metabólica	UPLC-MS GC x GC-TOF MS	Orešič y col., 2011
CSF 4 pool de muestras de pacientes con AD (n=27), MCI con progresión a AD (n=13), MCI sin progresión a AD (n=26), y controles (n=33).	Perfil de aminoácidos	MEKC-LIF	Samakashvili y col., 2011
Tejido post-mortem Pacientes con AD (n=8) y pacientes con esclerosis lateral amiotrófica (n=11).	Huella metabólica	NMR	Botosoa y col., 2012
CSF Pacientes con AD (n=23), MCI con progresión a AD (n=9), MCI sin progresión a AD (n=22), y controles (n=19). Validación (n=12).	Huella metabólica	CE-TOF MS	Ibáñez y col., 2012
CSF Pacientes con AD (n=10) y controles (n=10). **Plasma** Pacientes con AD (n=41), MCI (n=26) y controles (n=42).	Perfil de esteroides	HPLC-Q/MS LTQ Orbitrap	Sato y col., 2012
Tejido post-mortem de ratones Ratas control (n=3) y ratones transgénicos (APP/PS1) (n=3).	Perfil del metabolismo energético	GC-TOF MS	Trushina y col., 2012
Orina de rata Ratas control (n=4) frente a transgénicas (ratas APP/tau) a 4 (n=4), 10 (n=4) y 15 (n=3) meses de edad.	Huella metabólica	NMR	Fukuhara y col., 2013
Tejido post-mortem Pacientes con AD (n=15) y controles (n=15). Validación (n=60).	Huella metabólica	UPLC-Q/TOFMS	Graham y col., 2013
CSF Pacientes con AD (n=21), MCI con progresión a AD (n=12), MCI sin progresión a AD (n=21), y controles (n=21).	Huella metabólica	UPLC-TOF MS	Ibáñez y col., 2013
CSF Pacientes con AD (n=40), MCI (n=36) y controles (n=38).	Huella metabólica	HPLC-ECA[b]	Kaddura-Daouk y col., 2013
Tejido post-mortem Pacientes con AD (n=9) y controles (n=10).	Perfil de intermediarios del colesterol	HPLC-FIA MS NMR	Wisniewski y col., 2013

[a]FIA: Análisis por inyección en flujo; [b]ECA: detección electroquímica.

La **Segunda Part**e de esta Tesis contiene, además de una completa revisión bibliográfica de la AD, los trabajos de investigación llevados a cabo sobre la búsqueda de marcadores metabólicos involucrados en el progreso de AD en muestras de CSF humanas. Esta investigación se ha llevado a cabo en estrecha colaboración con el Dr. Ángel Cedazo-Mínguez del Instituto Karolinska (Estocolomo, Suecia), responsable de los estudios clínicos y de la toma de muestras de CSF de los pacientes que participaron en el estudio. Además, se ha colaborado con el Dr. Oliver Fiehn de la Universidad de Davis (California, EE.UU.) en el tratamiento bioinformático de los datos procedentes de los análisis mediante UPLC-MS. Esta última colaboración es fruto de una estancia pre-doctoral de cinco meses de duración durante el año 2012 en el laboratorio del Dr. Fiehn (http://fiehnlab.ucdavis.edu) financiada con una beca FPI del Ministerio de Economía y Competitividad.

En el **Capítulo 4** se muestra el potencial de la Metabolómica en el descubrimiento de biomarcadores relacionados con AD. Se trata de un trabajo de revisión bibliográfica en el cual se describen y discuten los trabajos publicados empleando estrategias metabolómicas no dirigidas en el estudio de AD (hasta Diciembre de 2012), así como las necesidades y futuras direcciones en esta área de investigación.

En los siguientes capítulos (**Capítulo 5, Capítulo 6 y Capítulo 7**) se han estudiado las diferencias metabólicas en muestras de CSF procedentes de sujetos con diferente estatus cognitivo. Concretamente se analizaron las muestras procedentes de individuos que recibieron un diagnóstico inicial de deterioro cognitivo subjetivo (SCI), MCI y AD. Tras un seguimiento clínico de dos años llevado a cabo en el Hospital Universitario Karolinska en Suecia, se observó que algunos de los pacientes, cuyo diagnóstico inicial era MCI, progresaron a AD, por lo que el grupo MCI se subdividió en aquellos individuos que presentaban un MCI estable (MCI-S) y aquellos que progresaron a AD (MCI-AD). Más concretamente, en el **Capítulo 5** se ha optimizado un método analítico enantioselectivo basado en MEKC-LIF quiral para la determinación de los perfiles aminoacídicos, así como de las diferencias en el contenido de D- y L-aa, entre los cuatro grupos de muestras de CSF (SCI, MCI-S, MCI-AD y AD). Por otro lado, en los dos últimos capítulos de esta Tesis (**Capítulo 6 y Capítulo 7**) se han comparado las huellas metabólicas de los individuos pertenecientes a los cuatro grupos

de muestras (SCI, MCI-S, MCI-AD y AD) mediante CE-TOF MS (**Capítulo 6**), y mediante HILIC/UPLC-TOF MS y RP/UPLC-TOF MS (**Capítulo 7**). En ambos capítulos se han aplicado técnicas estadísticas multivariantes con el fin de encontrar biomarcadores de detección temprana de AD y profundizar en los mecanismos moleculares que intervienen en el desarrollo de la enfermedad.

REFERENCIAS

Aftab M.F., Waraich R.S.; *Am J Neurosci* 3 (2012) 54-62.

Alzheimer A.; *Allgemeine Zeitschrift fur Psychiatrie und Psychisch-Gerichtlich Medizin* 64 (1907) 146-148.

Bartus R., Dean R., Beer B., Lippa A.; *Science* 217 (1982) 408-417.

Bartus R.T.; *Exp Neurol* 163 (2000) 495-529.

Bazenet C., Lovestone S.; *Biomark Med* 6 (2012) 441-454.

Beach T.G.; *J Alzheimers Dis* 33 (2013) S219-S233.

Blennow K., de Leon M.J., Zetterberg H.; *Lancet* 368 (2006) 387-403.

Botosoa E.P., Zhu M., Marbeuf-Gueye C., Triba M.N., Dutheil F., Duyckäerts C., Beaune P., Loriotb M.A., Le Moyec L.; *IRBM* 33 (2012) 281-286.

Cedazo-Minguez A., Winblad B.; *Exp Gerontol* (45) 2010 5-14.

Chong M.S., Sahadevan S.; *Lancet Neurol* 4 (2005) 576-579.

Dumas J.A., Newhouse P.A.; *Pharmacol Biochem Behav* 99 (2011) 254-261.

Francis P.T., Palmer A.M., Snape M., Wilcock G.K.; *J Neurol Neurosurg Psychiatr* 66 (1999) 137-147.

Fukuhara K., Ohno A., Ota Y., Senoo Y., Maekawa K., Okuda H., Kurihara M., Okuno A., Niida S., Saito Y., Takikawa O.; *J Clin Biochem Nutr* 52 (2013) 133-138.

Graham S.F., Chevallier O.P., Roberts D., Hölscher C., Elliott C.T., Green B.D.; *Anal Chem* 85 (2013) 1803-1811.

Greenberg N., Grassano A., Thambisetty M., Lovestone S., Legido-Quigley C.; *Electrophoresis* 30 (2009) 1235-1239.

Grimm M.O., Grösgen S., Riemenschneider M., Tanila H., Grimm H.S., Hartmann T.; *J Chromatogr A* 1218 (2011) 7713-7722.

Han X., Rozen S., Boyle S.H., Hellegers C., Cheng H., Burke J.R., Welsh-Bohmer K.A., Doraiswamy P.M., Kaddurah-Daouk R.; *PLoS One* 6 (2011) e21643.

Hardy J., Allsop D.; *Trends Pharmacol Sci* 12 (1991) 383-388.

Hassan-Smith G., Wallace G.R., Douglas M.R., Sinclair A.J.; *J Neuroimmunol* 248

(2012) 48-52.

Ibáñez C., Simó C., Martín-Álvarez P.J., Kivipelto M., Winblad B., Cedazo-Mínguez A., Cifuentes A.; *Anal Chem* 84 (2012) 8532-8540.

Ibáñez C., Simó C., Barupal D.K., Fiehn O., Kivipelto M., Cedazo-Mínguez A., Cifuentes A.; *J Chromatogr A* (2013). Sent.

Jellinger K.A., Bancher C.; *J Neural Transm Suppl* 54 (1998) 77-95.

Kaddurah-Daouk R., Rozen S., Matson W., Han X., Hulette C.M., Burke J.R., Doraiswamy P.M., Welsh-Bohmer K.A.; *Alzheimers Dement* 7 (2011) 309-317.

Kaddurah-Daouk R., Zhu H., Sharma S., Bogdanov M., Rozen S.G., Matson W., Oki N.O., Motsinger-Reif A.A., Churchill E., Lei Z., Appleby D., Kling M.A., Trojanowski J.Q., Doraiswamy P.M., Arnold S.E.; *Transl Psychiatry* 3 (201) e244.

Lista S., Faltraco F., Prvulovic D., Hampel H.; *Prog Neurobiol* 101-102 (2013) 1-17.

Manzano-Palomo S., de la Morena-Vicente M.A, Barquero M.S.; *Rev Neurol* 42 (2006) 350-353.

Mullane K., Williams M.; *Biochem Pharmacol* 85 (2013) 289-305.

Povova J., Ambroz P., Bar M., Pavukova V., Sery O., Tomaskova H., Janout V.; *Biomed Pap* 156 (2012) 108-114.

Reiman E.M., Jagust W.J.; *NeuroImage* 61 (2012) 505-516.

Román G., Pascual B.; *Arch Med Res* 43 (2012) 671-676.

Rylett R.J., Ball M.J., Colhuon E.H.; *Brain Res* 289 (1983) 169-175.

Salek R.M., Xia J., Innes A., Sweatman B.C., Adalbert R., Randle S., McGowan E., Emson P.C., Griffin J.L.; *Neurochem Int* 56 (2010) 937-947.

Samakashvili S., Ibáñez C., Simó C., Gil-Bea F.J., Winblad B., Cedazo-Mínguez A., Cifuentes A.; *Electrophoresis* 32 (2011) 2757-2764.

Sato Y., Suzuki I., Nakamura T., Bernier F., Aoshima K., Oda Y.; *J Lipid Res* 53 (2012) 567-576.

Shi M., Caudle W.M., Zhang J.; *Neurobiology Dis* 35 (2009) 157-164.

Sperling R.A., Aisen P.S., Beckett L.A., Bennett D.A., Craft S., Fagan A.M., Iwatsubo T., Jack C.R.Jr., Kaye J., Montine T.J., Park D.C., Reiman E.M., Rowe C.C., Siemers

E., Stern Y., Yaffe K., Carrillo M.C., Thies B., Morrison-Bogorad M., Wagster M.V., Phelps C.H.; *Alzheimers Dement* 7 (2011) 280-292.

Sutherland G.T., Janitz M., Kril J.J.; *J Neurochem* 116 (2011) 937-946.

Swerdlow R.H.; *Neurotox Res* 22 (2012) 182-194.

Tam J.H., Pasternak S.H.; *Can J Neurol Sci* 39 (2012) 286-298.

Tiraboschi P., Hansen L.A., Thal L.J., Corey-Bloom J.; *Neurology* 62 (2004) 1984-1989.

Trushina E., Nemutlu E., Zhang S., Christensen T., Camp J., Mesa J., Siddiqui A., Tamura Y., Sesaki H., Wengenack T.M., Dzeja P.P., Poduslo J.F.; *PLoS One* 7 (2012) e32737.

Tukiainen T., Tynkkynen T., Mäkinen V.P., Jylänki P., Kangas A., Hokkanen J., Vehtari A., Gröhn O., Hallikainen M., Soininen H., Kivipelto M., Groop P.H., Kaski K., Laatikainen R., Soininen P., Pirttilä T., Ala-Korpela M.; *Biochem Biophys Res Commun* 375 (2008) 356-361.

Von Bernhardi-Montgomery R.; *Rev chil neuro-psiquiatr* 43 (2005) 123-132.

Wenk G.L.; *J Clin Psychiatry* 64 (2003) 7-10.

White P., Hiley C.R., Goodhardt M.J., Carrasco L.H., Keet J.P., Williams I.E., Bowen D.M.; *Lancet* 1 (1977) 668-671.

Wimo A., Prince M. (Eds.) *Alzheimer's Disease International: World Alzheimer Report 2010: The Global Economic Impact of Dementia*. ADI, London, (2010).

Wisniewski T., Newman K., Javitt N.B.; *J Alzheimers Dis* 33 (2013) 881-888.

4.2.   CAPÍTULO 4 (Chapter 4)

Metabolomics in Alzheimer's Disease Research

Metabolomics in Alzheimer's Disease Research

Clara Ibáñez, Carolina Simó*, Alejandro Cifuentes

Laboratory of Foodomics, CIAL (CSIC), Nicolás Cabrera 9, 28049 Madrid, Spain.

CORRESPONDENCE: Dr. Carolina Simó. **Address:** Laboratory of Foodomics, CIAL (CSIC), Nicolás Cabrera 9, 28049 Madrid, Spain; **E-mail**: c.simo@csic.es; **Tel.** (+34) 910017947; **Fax** (+34) 910017905.

Abbreviations: **Aβ**, amyloid beta ; **AD**, Alzheimer's disease; **ApoE**, apolipoprotein E; **APP**, amyloid precursor protein ; **CNS**, central nervous system; **CSF**, cerebrospinal fluid; **DFI**, direct flow injection; **ECA**, electrochemical array detection; **LDA**, linear discriminant analysis; **MCI**, mild cognitive impairment; **MDMS-SL**, multi-dimensional mass spectrometry-based shotgun lipidomics; **Mo**, monomeric-enriched fraction; **MRI**, magnetic resonance imaging; **NFT**, neurofibrillary tangles; **PCA**, principal component analyisis; **PET**, positron emission tomography; **Po**, polymeric-enriched fraction; **P-tau**, phosphor-tau; **Q**, quadrupole; **QqQ**, triple quadrupole; **SPECT**, single photon emission computed tomography; **T-tau**, total tau; **UPLC**, ultra performance liquid chromatography.

Abstract

Alzheimer's disease (AD) is a neurodegenerative multifactorial disease whose cause is still unknown. The majority of AD biochemical markers currently available have been developed as an extension of targeted physiological studies on the basis of the "amyloid hypothesis". The potential of Metabolomics for the discovery of novel biomarkers and elucidation of new biochemical pathways modified in the progression of AD is highlighted in this review work. A variety of non-targeted metabolomic approaches for the discrimination between healthy subjects and AD patients are described. Moreover, the feasibility of Metabolomics to predict progression to AD in individuals with mild cognitive impairment (MCI) is also presented.

1. Introduction to Alzheimer's Disease.

The incidence of many diseases increases rapidly with aging. Alzheimer's disease (AD) is the most prevalent cause of dementia among older people [1]. It is an incurable, degenerative, and terminal multifactorial disease. Although the initiating events are still unknown [2], AD seems to result from a combination of genetic, environmental and lifestyle risk factors [3]. One of the hallmarks of AD is the observation of amyloid plaques (or senile plaques) and neurofibrillary tangles (NFT). Amyloid plaques are produced by the accumulation of amyloid beta (Aβ) peptides. The "amyloid hypothesis" assigns a critical role to abnormal processing of amyloid precursor protein (APP), which is sequentially cleaved by α-, β- and γ-secretase originating neurotoxic soluble Aβ peptides that aggregate in oligomers to form these plaques. On the other hand, NFT are intracellular filamentous aggregates of the microtubule-associated protein tau. In its hyperphosphorylated status, tau protein detaches from the microtubules and, consequently, the microtubules fall apart and tau tends to aggregate inducing breaks in the microtubular tracks and neuronal death. It has been estimated that AD process begins more than 20 years before the clinical onset of dementia [4]. AD is a very long and progressive disorder and detection of transition points for single patients has not been elucidated yet. AD is preceded by a mild cognitive impairment (MCI) state followed by dementia [5]. However, up to date, there is no clinical method to determine which MCI cases will progress to AD except for a long clinical follow-up period. MCI is a syndrome defined as cognitive decline greater than expected for an individual's age and education level but that does not interfere notably with activities of daily life [6, 7]. In Figure

1, the existing diagnostic flowchart used to classify the subtypes of MCI is presented [8]. In approximately 80% of cases, MCI progresses to dementia when these subjects are followed for up to 6 years [9]. On the other hand, it has been suggested that there is a preclinical phase (presymptomatic) which progresses to MCI [10]. Even though, the early molecular events that occur in AD are not clear yet. In a world of aging population it has been estimated that 65.7 million people would suffer dementia in 2030 and 115.4 million in 2050 [11]. As public awareness of AD increases, the need for methodologies for early diagnosis is becoming imperative.

2. Current biomarkers in diagnosis and prediction of progression.

Advanced medical brain imaging techniques have greatly contributed to the scientific understanding of AD and currently, they provide comprehensive and non-invasive information of cerebral morphology and metabolism and may help to the detection and tracking of AD. The best established brain imaging technologies include structural magnetic resonance imaging (MRI), single photon emission computed tomography (SPECT) and positron emission tomography (PET) [12, 13]. As imaging advances are being presented, parallel work is being carried out to identify reliable and valid markers in biofluids indicative of AD pathology [14-17]. Most available biochemical markers have been developed as an extension of targeted physiological studies. On the basis of the "amyloid hypothesis" [18], the focus on amyloid biomarkers has been the measurement of $A\beta_{1-42}$, the least soluble of the known $A\beta$ isoforms. Low $A\beta_{1-42}$ peptide concentration in CSF as well as increased total tau (T-tau) and phosphor-tau (P-tau) have been reported as an indicator of amyloid deposition in patients diagnosed as converting from MCI to AD [15, 19, 20]. On the other hand, measurements of AD biomarkers have shown large variability inter-laboratories due to a variety of reasons [21].

Identifying soon the occurrence and/or progression of AD will increase our knowledge on potential AD risk factors, advance on efficient drug discovery research and cognitive stimulation programs. There is no established method to predict progression to AD in individuals with MCI. Thus, an imperative need arises to identify specifically those MCI patients who will later progress to AD. The combination of T-tau, P-tau181 and Aβ42 levels in MCI patients was correlated with their later development of AD.in a 4/6-year follow-up study. These three molecules were found to be robust and independent biomarkers for AD development in patients with MCI. Values of 95% sensitivity and 83%

specificity for detection of incipient AD in patients with MCI were reported [22]. However, as indicated by the authors, the specificity of these
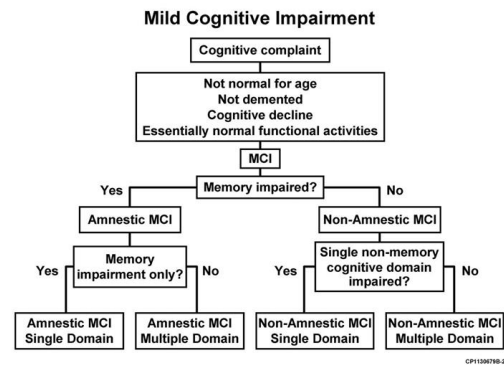
Figure 1. Flowchart for diagnosis of mild cognitive impairment (MCI) subtypes (from Petersen et al. [8]).

CSF biomarkers might be increased since some of the stable MCI cases with pathological CSF might still develop AD later on. Meanwhile, there is a need for the development of robust tools to identify presymptomatic individuals with a risk for developing the disease. This fact will help clinicians in selecting those patients who would be most likely to benefit from an early disease-modifying treatment against AD and to monitor responses to new therapies. Anyhow, a unique biomarker unequivocally associated with AD causality has not been found yet. Increased CSF T-tau is observed in neuronal degeneration which is not only found in AD. Aβ deposition has also been observed in normal ageing [23] and thus, reduced soluble CSF Aβ42 is not specific to AD. The combination of the analysis of specific protein levels and brain imaging might increase AD diagnose up to 90% [24] increasing enormously the cost per patient (ca. 6,000-10,000 euros approximately). In any case, today AD can be definitively diagnosed only after death through an examination of brain tissue and pathology in an autopsy.

With the introduction of advanced NMR and MS instruments, the search of biomarkers implicated in the pathogenesis of AD is moving away from traditional targeted to non-biased profiling approaches of human fluids in an attempt to discover novel biomarkers [25, 26]. Among the "omic" platforms used to perform research into the biomarker discovery in AD field following an unbiased approach, Proteomics has attained important consideration. The potential of Proteomics for biomarker discovery in AD research is evidenced by the recent review works published in this subject [27, 28].

Metabolomics is one of the newest "omics" disciplines. It is focused on the analysis of the complete set of small-molecules synthesized by a

biological system which participate in metabolic reactions required for growth, maintenance and normal function of an organism. Metabolomics offers complementary insight into the full complexity of the disease phenotype [29]. The practice of Metabolomics in AD research should be regarded as exploratory and meant for hypothesis and model generation, rather than for hypothesis confirmation and model validation. This novel approach is based on the fact that pathologically distinct diseases are characterized by unique metabolite profiles [30, 31]. At a time in which the causes of neurological diseases are still not clear, Metabolomics opens new frontiers as hundreds of metabolites can be measured rapidly and simultaneously [32]. In next sections an up-to-date overview on existing Metabolomics-based applications for the study of AD is presented.

3. The practice of Metabolomics in AD research.

In the last decade, Metabolomics has been demonstrated to be a promising discipline able to generate disease-specific metabolite signatures unique to individuals [33]. It is concerned with identification and quantification of small molecules (<1500 Da) in a high-throughput manner. One of the main challenges in Metabolomics is the complete analysis of entire metabolome in biological systems since metabolites have very different molecular structures (lipids, carbohydrates, nucleic acids, amino acids, organic acids, steroids, peptides and others) and are present in biofluids and tissues in a wide dynamic range of concentrations. Two analytical platforms are by far the most predominantly used for metabolomic analyses: MS- and NMR-based analytical platforms. NMR permits the analysis of complex mixtures of metabolites with little or no sample preparation in a rapid and non-destructive way [34]. MS technique is highly sensitive and provides spectral information for the identification process of metabolites [35]. The use of ultra-high resolution mass spectrometers is essential in metabolomics field to obtain accurate mass measurements for the determination of elemental compositions of metabolites, and to carry out their tentative identification with the help of metabolite databases [36]. Metabolomic analysis, using either NMR spectroscopy or MS can be performed using biofluids such as urine, blood or cerebrospinal fluid (CSF) and may permit the identification of disease specific metabolite signatures which may be useful as disease biomarkers [33, 37, 38]. Two different analytical approaches can be followed in non-targeted Metabolomics. "Metabolic profiling" is referred to analyze a class of metabolites (usually chemically related metabolites what facilitate simultaneous analysis, e.g., amino acids,

alcohols, carbohydrates, etc), and "metabolic fingerprinting" has been proposed as a means of analyzing the total set of metabolites for rapid sample classification via multivariate statistical analysis. Metabolic signals used as dominant classifiers will be subsequently identified to define potential biomarkers related to prognosis or diagnosis of a disease, or to drug effect. An overview of a typical Metabolomics workflow in AD research is given in Figure 2. Given the novelty of this approach, applications to neurological diseases are just beginning [32]. As well as for other diseases [39], metabolomic study of biological fluids and tissues will also provide insights for a better understanding of molecular, structural, and functional changes that are causally related to the onset of age-related diseases. Although Metabolomics is still in the early stages of growth in AD investigation, a number of research works have been presented in the last years.

3.1. CSF Metabolomics in AD. Metabolomics approaches described in literature are mostly based on the analysis of biofluids or animal models. CSF represents the most direct biofluid to study the biochemical changes occurring in the central nervous system (CNS). Metabolic profiling of CSF biofluid has been strongly considered since it is the only body fluid in direct contact with the extracellular space of the brain and thus biochemical changes due to pathological brain-processes are more probable to be reflected in CSF than in other body-fluids. Lumbar puncture is the procedure required to collect CSF samples. At present the updated human CSF metabolome information obtained from five analytical platforms (NMR, GC-MS, LC-MS, DFI-MS/MS, ICP-MS) contains 468 identified metabolites [40]. With the continuous advances in the analytical technologies for Metabolomics (better instrumentation sensitivity, enhanced resolution, improved software, etc.) the number of studied metabolites in CSF will be continuously increasing. One of the major advantages of NMR is that minimal sample processing is required, although at the expenses of sensitivity. In a very fist work, Nicoli et al. demonstrated the applicability of NMR as analytical platform for the study of CSF metabolic profiles in AD [41]. A group of 12 patients diagnosed with neurodegenerative dementia (ten had probable AD) were compared with a control group of 17 patients which did not suffer from any inflammatory, degenerative or tumoral disease of the CNS. Increased concentrations of lactate, pyruvate, alanine, lysine, valine, leucine/isoleucine, tyrosine and glutamine were observed in patients with AD dementia. This effect

was related to a brain oxidative metabolism impairment already described in AD. Using similar methodology, NMR-based metabolomic approach normal subjects (n=15) were studied using this metabolomic approach. The use of ECA detection in Metabolomics is less usual due to its lower
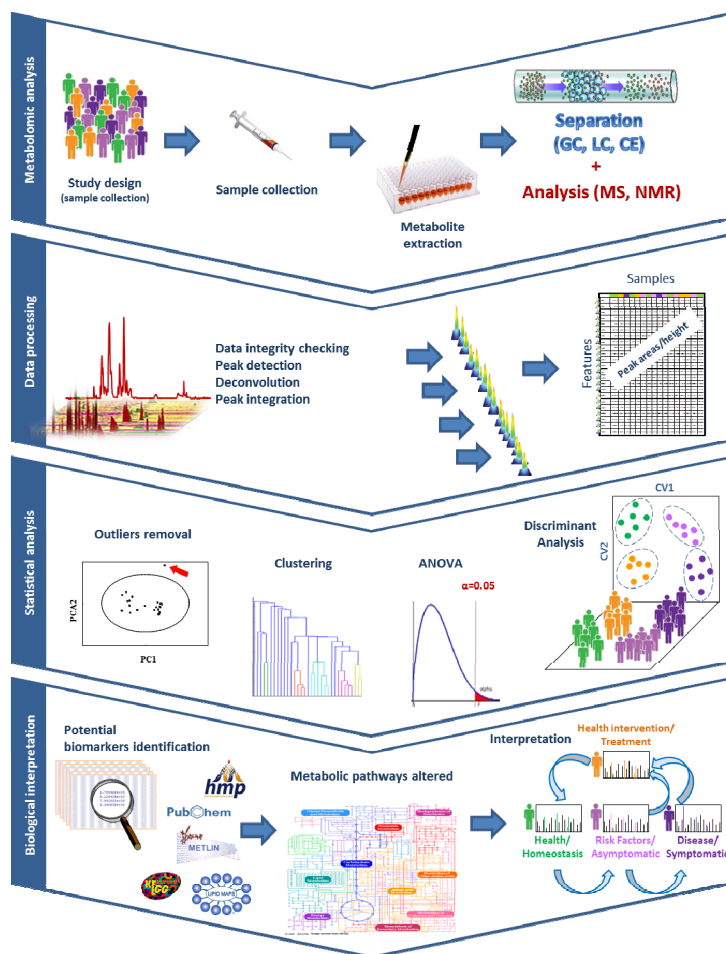
Figure 2. Metabolomics workflow in AD research.

was also applied to CSF samples obtained from AD patient's autopsy and from control subjects without any neurological disease [42]. Metabolomic analysis clearly distinguished AD patients from control subjects mainly in the delta 2.4-2.9 region of the NMR spectra. Resonances attributable to citrate, aspartate, N-acetyl aspartate, methionine and glutamate were assigned in this region of the spectra. A subsequent unsupervised PCA model separated the two studied groups, and in a later univariate statistical analysis it was observed that citrate level was the principal marker among the studied groups.

With the aim of understanding AD mechanisms and identifying potential candidate markers of disease, LC with electrochemical array detection (LC-ECA) was used for metabolic profiling of postmortem ventricular CSF [43]. Biochemical changes that discriminated autopsy-confirmed diagnosed AD subjects (n=15) from cognitively

capability to give structural information compared to NMR or MS. On the other hand, ECA detection permits the detection of metabolites on the basis of their oxidation-reduction potentials with simpler and lower-cost instrumentation. Metabolic extracts from CSF samples were obtained after protein precipitation with cold acidic acetonitrile. After metabolite profiling by LC-ECA of CSF extracts multivariate and univariate statistical analysis were performed. It was observed that norepinephrine was significantly decreased in patients with AD. Other differences in metabolite content among the two groups of samples revealed modifications in the tyrosine, tryptophan, purine, and tocopherol pathways. Although the study of postmortem ventricular CSF is a good starting point to develop novel strategies based on metabolomics approaches, it presents several disadvantages. One of the main disadvantages is the influence of the death process on CSF metabolome. On the other

hand, CSF metabolomic studies of early stage disease might have greater relevance to clinical biomarker development. Recently, a novel MECK-LIF method to study the relationship between enantiomeric composition of amino acids in human CSF samples and AD was carried out by Samakashavili et al. [44]. With the basis that free amino acids are important in neurotransmission, receptor function and are implicated in neurotoxicity, change in the metabolism of free amino acids was studied as potential early indicator of neurodegeneration in AD. When CSF samples were analyzed by MECK-LIF, significant lower levels of L-Arginine, L-Lysine, L-Glutamic acid and L-Aspartic acid and significant higher level of GABA were observed in subjects who presented AD. Although sensitive is challenging using MECK due to the low injected sample volume, LIF detection offers extremely low detection limits. On the other hand, structural information provided by LIF detection is limited. Today, metabolomics approaches are dominated by mass spectrometry. One of the main advantages of MS-based metabolomics approaches is that they offer an excellent combination of sensitivity and

developed a new chromatographic method for profiling of hydrophilic metabolites in CSF samples from AD diagnosed subjects (n=18) and healthy controls (n=18). After CSF deproteinization with a 5 kDa centrifuge filter, metabolite purification was carried out by SPE using a polymeric sorbent. Finally, analysis of CSF extracts was carried out by nanoLC-Q/TOF in the positive ionization mode (Figure 3). Extensive optimization of stationary phases for separation of weak and strong cationic metabolites was also presented. By using multivariate analysis satisfactory separation was obtained among two groups of subjects using non-supervised PCA statistics. Since no single analytical method can cover the chemical diversity of the entire metabolome, the use of multiplatform-based approaches are increasingly being used in Metabolomics. In order to cover a wider spectrum of metabolite polarities, a doubled analytical platform for the analysis of diagnosed AD subjects and healthy controls has been proposed [46]. GC-MS and LC-MS/MS techniques were used for the analysis of 130 deproteinized CSF samples. AD patients divided in two sub-groups (light/moderate
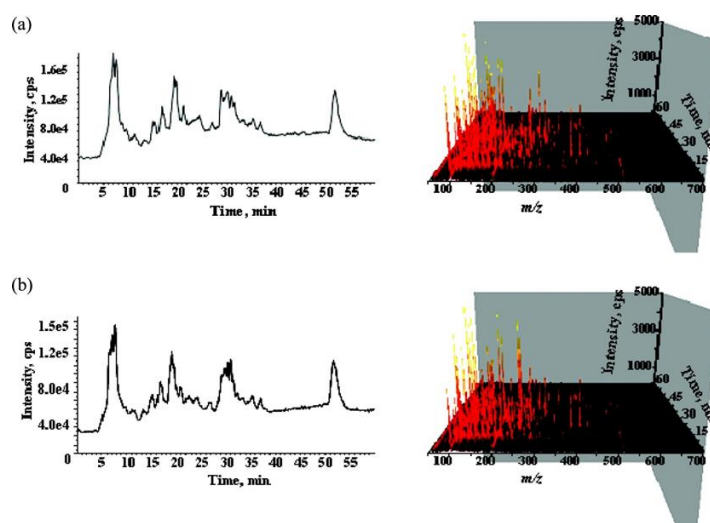
Figure 3. NanoLC-Q/TOF total ion chromatograms of CSF samples from AD (a) and healthy control (b) subjects (from Myint et al. [45]).

selectivity. Indeed, the number of MS-based studies now exceeds that of NMR-based metabolomics studies. Conversely, MS-based approaches usually require extensive sample clean-up prior to analysis. MS is often preceded by chromatographic (LC, GC) or electrophoretic separations (CE). Analysis of polar metabolites is specially challenging in Metabolomics when working with LC. For this reason, Myint et al. [45]

AD (n= 53) and moderate/strong AD (n= 26)), and healthy control subjects (n=51) were studied. Unsupervised PCA, supervised OPLS-DA multivariate analysis as well as univariate analysis were carried out to study metabolic differences to distinguish between the different diagnostic groups. It was observed that changes in brain metabolism in AD could be detected in CSF samples. Different combinations (pairs, triplets,

quartets, quintets) of specific metabolites (uridine, cortisol, cysteine, dopamine, methionine, phenylalanine, serine, among others) gave good prediction values with sensitivity and specificity values above 80%.

Monitoring the evolution and progression of AD is particularly challenging since the cause and progression of AD are still not well understood. MCI represents the earliest clinical state of AD and thus, it has become a focus in the search of biomarkers using metabolomic approaches. Recently, Ibañez et al. [47] proposed a new CE-MS-based metabolomic approach to study AD progression from early stages. As well as in other biomarker discovery-related works, CSF samples were considered since they offer the highest yield of biomarkers related to AD because of its direct contact with the brain. In a 2-years follow-up study, CSF samples from 85 subjects were obtained from control non-demented subjects (n=19), with MCI patients which remained stable after the 2-years follow-up period (n=22), MCI subjects which progressed to AD (n=9) and diagnosed AD patients (n=23). Prior to its analysis, CSF samples were ultrafiltrated by using a 3 kDa cut-off filter. Metabolic CSF fractions (compounds < 3 kDa) were then analyzed by CE-MS. CE-MS analytical platform, is particularly suited for the rapid separation of ionic, weakly ionic and/or highly polar metabolites with very high resolution using extremely small reagents and sample volumes (several nanoL). In order to develop the predictive model, a linear discriminant analysis (LDA) with the 160 metabolic signals detected in the entire set of samples was applied to separate the four groups of patients. From this analysis, a 90% of the subjects were correctly diagnosed using 10 metabolites as classifiers. It was also observed that the highest metabolic similarity was achieved between the subjects who did not develop AD (healthy and MCI patients which remained stable). These two diagnostic groups were considered the same diagnostic group to develop a second predictive model using LDA. Certainly, when these two groups were merged, higher predictive diagnostic power was obtained reaching 97% using the levels of 14 metabolites with choline leading in importance and encompassing some metabolites highly related with energy provision, namely creatine and carnitine, with consequent lower levels in AD patients. When 12 blind CSF samples were subjected to classification using the later prediction model, only two subjects were misdiagnosed of which one is not completely clear yet (i.e., a young individual who can evolve to AD) (Figure 4).

3.2. Serum/plasma Metabolomics in AD.

CSF is an attractive sample for AD research, however, CSF collection still remains an invasive

procedure requiring a lumbar puncture that can lead to post-lumbar puncture headache. An ideal biomarker should show up in a simple non-expensive blood/urine test, thus non-invasive. Collection of biofluids such as urine [48] or minimally invasive collection of blood [49, 50] are generally preferred for a routine implementation in clinical practice. Ease of collection also enables serial sampling for evaluation of disease-modifying treatments and assessment of prognosis. With the idea of developing a non-invasive assay able to both diagnose and monitor AD, several works have been described the analysis of less invasive biofluids, such as urine or plasma [51, 52]. Thus, Greenberg et al. [52] followed a metabolomic-based strategy to study differences in plasma samples from subjects with different cognitive status related to AD progression. However, plasma metabolic differences obtained by UPLC-Q/TOF MS between AD (n= 16), MCI (n= 12) and normal elderly (n= 10) subjects were not enough to differentiate among groups. This result highlights the importance of a sufficient number of samples. Inter- and intra-subject biological variation must carefully be considered in Metabolomics. Factors like gender, age and diet have been proven to have effect on biofluids metabolic profiles [53] and thus, a large number of samples should be considered to overcome biological variation.
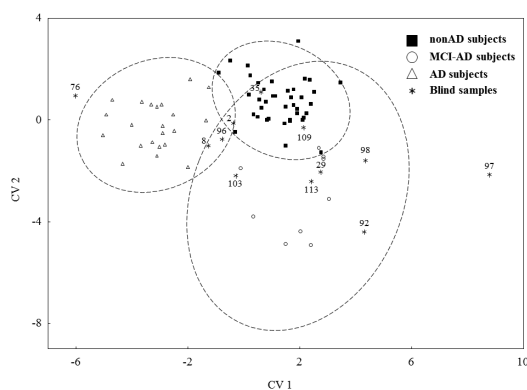
Figure 4. Plot for the classification of CSF samples from individuals at different stages of AD progression. The plane is defined by the two canonical variables (CV1 and CV2) obtained by CVA after CE-MS metabolomics of CSF (from Ibañez et al., [47]).

With the aim of studying AD progression using a Metabolomic-based approach, Oresic et al. [54] studied plasma metabolome alterations at different stages of the disease in a large cohort of patients in a 1-4-year follow-up study. Thus, plasma samples from healthy control subjects (n=46), AD diagnosed patients (n=37), MCI subjects whose status remained stable or improved in the follow-

up (n=91) and MCI subjects which progressed to AD in the follow-up period of time (n=52) were analyzed. In order to have a broad analytical coverage of plasma metabolome, two analytical platforms were used for this purpose. Bidimensional GCxGC-TOF MS was selected for the analysis of amino acids, free fatty acids, ketoacids, organic acids, sterols and sugars. On the other hand, for lipidome analysis, UPLC-Q/TOF MS platform was used. Sample treatment was adapted in each case to obtain a better analytical performance. Thus, prior to GCxGC-TOF MS analysis, a methanol extraction was carried out followed by a derivatization process to convert metabolites into their methoxime and trimethylsilyl derivatives. On the other hand, plasma lipidic fraction analyzed by UPLC-TOF MS was obtained after a simple extraction procedure using a chloroform/methanol mixture. After metabolomic analyses, 139 molecular lipids and 544 polar metabolites were measured. Combination of these two analytical platforms allowed the determination of intermediates in a variety of metabolic pathways, namely, lipid metabolism, energy metabolism (tricarboxylic acid cycle, gluconeogenesis, ketogenesis) and nitrogen metabolism. AD patients were characterized by diminished ether phospholipids, phosphatidylcholines, sphingomyelins and sterols. Prediction of progression to AD was based on a group of three metabolites (2, 4-dihydroxybutanoic acid, an unidentified carboxylic acid and phosphatidylcholine 16:0/16:0) (Figure 5). Major contributing metabolite in the marker panel separating the two MCI subgroups was 2, 4-dihydroxybutanoic acid which was found to be upregulated in MCI patients which progressed to AD, indicating potential involvement of hypoxia in the early AD pathogenesis.

A non-targeted metabolomic approach of plasma samples was also proposed by Li et al. [55]. Thus, metabolomic analysis of plasma samples from AD subjects (n=20) and healthy controls (n=20) was performed after protein precipitation with acetonitrile. Analysis of metabolic extracts was performed on UPLC-QqQ MS in the positive and negative ion mode. Based on PCA statistics, distinction of AD patient and control group were obtained observing different levels in plasma of lysophosphatidylcholines, sphingosine and
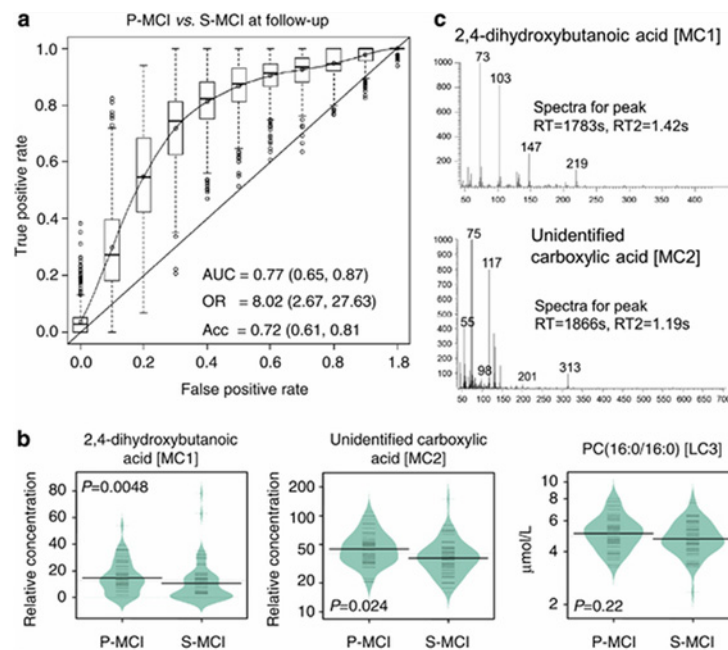
Figure 5. Feasibility of predicting AD, based on concentrations of three metabolites in subjects at baseline, who were diagnosed with mild cognitive impairment (MCI). (a) Characteristics of the model: Acc, classification accuracy; AUC, area under the receiver operating characteristic; OR, odds ratio. (b) Beanplots of the three metabolites included in the model. (c) GCxGC–TOF MS spectra of the two metabolites included in the model.

tryptophan. These results suggested perturbations of lecithin, amino acid and phospholipids metabolism in AD patients. Although a non-targeted metabolomic approach was presented by these authors, lipid dysregulation was mainly found to be related to AD. Similar approach was followed by Gonzalez-Dominguez et al. [56]. After applying a non-targeted metabolic approach of serum samples from AD patients (n=22) and healthy controls (n=18), a group of phosphatidylcholines were the main metabolic classifiers which permitted the differentiation between groups. Indeed, functioning of membrane proteins of neurons depend on membrane potential, receptor occupancy, phosphorylation-dependent intermolecular associations, and protein and lipid composition. Hypothesis that AD can be caused or induced by nonamyloid factors (calcium dysregulation, proteolysis failure, altered cell signaling, oxidative stress and inflammation) is currently under consideration in AD research [57]. Important research is being carried out to find out the role of lipids in AD [58-60]. Soluble beta-amyloid oligomers, considered proximate effectors of the neurotransmission injury, have been proposed to interact with the synaptic membrane causing synaptic injury and neuronal death. This neurodegeneration due to impairment of synaptic functions uncovers cell membranes as new target to trigger in AD research. In a recent work it was revised the importance of an adequate membrane lipid content as key for AD prevention with a special emphasis on n-3 polyunsaturated acids as major constituents of neuronal lipids [61]. Special emphasis to the analysis of lipids using non-targeted lipidomics approaches will be given in this review. Several authors have been addressed the lipidomic analysis in plasma/serum samples [62]. These results will be discussed below in section 3.4.

3.3. Brain tissue Metabolomics in AD.

The availability of brain tissue from AD studies in humans is practically not viable. Due to this, "omic" studies have been done using post-mortem brain tissue samples, however, the main limitation here is that disease is at its end-stage. To circumvent these limitations a variety of animal models are now being used through various genetic, biochemical or dietary manipulations [63,64] with the aim to elucidate disease mechanisms and thereby enable effective approaches to treat, and in the end, prevent the progression of AD. With the purpose of deepening on the knowledge of early mechanisms underlying AD and identifying associated biomarkers, Trushina et al. [65] focused their work in the impact of the mitochondrial dysfunction in early AD. Thus, brain tissue from three different mouse

models with familial AD mutations on mitochondrial functions were analyzed by GC-MS and LC-MS. Metabolomic profiling revealed mutation-specific changes in the levels of metabolites reflecting altered energy metabolism. Besides, alterations in nucleotide, Krebs cycle, energy transfer, carbohydrate, neurotransmitter, and amino acid metabolic pathways were observed.

Due to the lack of effective long-term medication for the treatment of patients with AD, parallel strategies are currently under development in order to avoid progressive neurodegeneration. Thus, the potential role of nutrition in the development of AD is currently under study [66-68]. As well as it has been suggested a protective effect of several bioactive compounds from diet on cardiovascular diseases [69, 70], it has been suggested that antioxidants from food may be associated with a lower incidence of AD [71] following the vascular component of AD pathology. Recent studies are focused on the use of bioactive compounds from plants as novel therapeutic agents to decrease cognitive deterioration in AD and other forms of neurodegenerative disorders [72, 73]. In a recent paper, Wang et al. [74] studied the effect of plant-derived polyphenolic compounds on cognitive behavioral in animal models of AD (namely, Tg2576 AD transgenic mice). Following the hypothesis that amyloid (or its precursors) is critical in the initiation of the disease, Tg2576 mice are being used to test therapeutic agents that may have utility in patients with AD. Tg2576 mouse expresses human APP at a level more than six-fold higher than endogenous murine APP. Usually this mouse has a fivefold increase in the concentration of $A\beta_{40}$ and a 14-fold increase in that of $A\beta_{42}$. Dietary supplementation with two different polyphenolic-enriched fractions from grape, namely, proanthocyanidin/catechin/epicatechin monomeric-enriched fraction (Mo) and a polymeric-enriched fraction (Po) were tested in Tg2576 mice [74]. After dietary supplementation during five months it was observed that supplementation with Mo significantly improved the cognitive behavioral performance of Tg2576 mice. In contrast, Po treatment did not lead to detectable improvements. Bioavailability of polyphenols and their metabolites in the brain was also studied. In order to study metabolic changes in brain, administration of Mo and Po fractions to Sprague-Dawley rats over a 10 day period was carried out. After SPE extraction Mo and Po metabolites from brain tissues extracts were analyzed by ESI-QqQ MS. It was found that only Mo metabolites selectively reached and accumulate in the brain. Free catechin and epicatechin and higher contents of catechin and epicatechin phase II metabolites (both O-β-

glucuronides and O-Me-β-glucuronides) were found. These results indicated a preferential accumulation or metabolism of methylated metabolites from the blood, or potential differences in the kinetics of transferring methylated metabolites across the blood-brain barrier. In this work authors suggested that studied polyphenols were involved in neuronal function restoring associated with learning and memory in the AD brain.

3.4. Lipidomics: A non-targeted profiling Metabolomics case in AD research.

Due to the vast number and complex structural diversity of lipids compared to most water soluble metabolites, non-targeted lipidome analysis (Lipidomics) is becoming increasingly important and extensive field inside Metabolomics [75-77]. Analyzing complex lipids in a mixture is a challenging task since they present high structural diversity due to the complex combination of hydrophobic acyl chain molecular species and hydrophilic functional groups. Lipidomics may also be used for the broad study of a biofluid/tissue to discover novel lipid biomarker candidates of a certain disease. As well as in other areas of Metabolomics, lipid biomarkers associated with specific intermediate phenotypes or disease endpoints are of great interest [78-80].

ApoE (apolipoprotein E) is one of the key lipoproteins involved in the regulation on the metabolism of lipids by directing their transport, delivery, and distribution from one tissue or cell type to another. There is strong evidence that ApoE isoforms differentially modulate amyloid-β metabolism and accumulation. Among the three AapoE isoforms (ApoE2, E3 and E4) and the six possible genotypes (ApoE2/2, E2/3, E2/4, E3/3, E3/4 and E4/4) determined by three alleles (ε3, ε2, and ε4), individuals with one or two copies of ApoE4 have a higher risk of developing AD ("sporadic" AD, which accounts for the majority of AD cases), compared with carriers of other ApoE isoforms which might even present protective properties [81]. However, the details of this process as well as the role that ApoE plays in non-Aβ-mediated mechanisms in AD pathogenesis remain to be fully clarified [82, 83]. As ApoE plays a prominent role in the transport and metabolism of several classes of lipids, the study of alterations in the lipid profiles of certain tissues/biofluids using lipidomic-based approaches is providing new insights into the biochemical mechanisms altered in AD patients [84, 85]. As in Metabolomics, LC-MS is being used for the global profiling of lipids in biological systems [86-88]. Following the hypothesis that lipid dysfunction plays an important role in AD pathogenesis, unbiased analysis using LC-LTQ Orbitrap MS has been recently presented for the comparison of

sterol profiles between AD (n=10) and elderly healthy controls (n=10) [62]. For sterol-related metabolite extraction from plasma, liquid-liquid extraction with hexane was followed. When univariate analysis was performed, significant changed peaks (fold change less than 0.5 or more than 2) among two groups of samples were extensively studied. After metabolite identification process, it was found that desmoterol and desmoterol/cholesterol ratio were significantly decreased in plasma from AD patients. Taking into account this preliminary finding, target analysis of desmoterol and cholesterol by GC-MS was carried out in CSF samples to study the correlation between plasma and CSF samples. Although desmoterol and cholesterol analysis were not significantly different between AD and non-AD subjects in CSF, interestingly, a good correlation between CSF and plasma was observed considering desmoterol/cholesterol ratio. Later on, specificity of these biomarkers was studied through their target analysis in plasma and CSF from patients with other neurological diseases (namely, schizophrenia and Parkinson's disease). It was reported that desmoterol was only decreased in AD patients. Further study was carried out in order to confirm desmoterol as a potential biomarker of AD. For this purpose, it was determined in a higher number of plasma samples from AD (n= 41), MCI (n= 26) and healthy elderly controls (n= 42). Finally it was confirmed that desmoterol and desmoterol/cholesterol ratio was significantly decreased in AD and MCI patients compared with healthy elderly subjects.

As already presented in section 3.2, successful application of lipid profiling using UPLC-Q/TOF MS coupling was recently published by Oresic et al. to study plasma metabolome changes in the progression of AD [54]. UPLC-MS based approaches have been used in Metabolomics to reduce the coelution of lipids since UPLC allows highly efficient separation in a very short analysis time [87]. However, with the introduction of high mass accuracy and high sensitivity MS, lipid extracts obtained from biological samples directly infused to a tandem mass spectrometer (without prior chromatographic separation) is becoming increasingly important in what is generally called "Shotgun Lipidomics". Alterations in plasmalogen (subclass of ethanolamine glycerophospholipids characterized by the presence of a vinyl ether substituent at the sn-1 position of the glycerol backbone) profiles in different regions of post-mortem human brain tissue was examined by using direct infusion negative-ion ESI-QqQ MS [89]. An acute decrease in plasmalogen content in white matter was observed in subjects at a very early stage of AD. This deficiency is likely to be a result of myelin sheath defects and is likely to result in axonal dysfunction, and thus may directly

contribute to the dementia in AD patients in this early stage. It was also found that the deficiency in plasmalogen in gray matter from several cerebral regions became more severe as AD progressed.

Multi-dimensional mass spectrometry-based shotgun lipidomics (MDMS-SL) has also demonstrated to be a successful approach in AD research [90-92]. In MDMS-SL the first dimension is composed by the molecular ions in m/z values, while the second dimension is comprised of the mass corresponding to the neutrally lost fragments or the monitored fragment ions in m/z values. The cross peaks of a given primary molecular ion in the first dimension with the second dimension represent the fragments of a given molecular ion. The major difficulty of this approach is the accurate interpretation of spectra. On the other hand, the main drawback of this methodology is ion suppression that can be partially avoided with exhaustive sample purification. Han et al. demonstrated the capabilities of MDMS-SL for the determination of alterations in the lipid profiles of post-mortem brain tissues from AD subjects [90]. Following this lipidomic approach sulfatide, galactocerebroside, sphingomyelin and ceramide content of different brain regions was associated to AD pathology as well as the stage of dementia in AD patients. 22 subjects whose cognitive status at time of death varied from no dementia to very severe dementia were studied for this purpose. Lipidic fractions were obtained by using chloroform/methanol as extraction solvents and analyzed by ESI-QqQ MS in the negative ion mode. Main observation was the depletion of the sulfatide content up to 92% in gray matter and up to 58% in white matter of all examined brain regions in the earliest stage of clinically identifiable AD patients. On the contrary, ceramides (potential degradation products of sulfatide) content was elevated at the stage of very mild dementia associated to AD. These findings were further investigated to elucidate the relationship between sulfatide depletion and the role of ApoE in AD patients. In this sense, subsequent studies aimed to identify the biochemical mechanisms underlying sulfatide depletion in early AD. Association of ApoE with sulfatide transport was suggested, being sulfatide homeostasis in the nervous system influenced through lipoprotein metabolism pathways. Thus, alterations in ApoE-mediated sulfatide trafficking can lead to sulfatide depletion in the brain [91].

Following the above-mentioned interest in the search of biomarkers in non-invasive biological samples, analysis of plasma from AD patients (n=26) and cognitively normal controls (n=26) was conducted following a similar MDMS-SL approach [92] with the objective to study alterations in plasma lipidome of AD subjects. Using this approach, over 800 lipid species (choline glycerophospholipids, lysophosphatidylcholines, ethanolamine glycerophospholipids, phosphatidylinositols, sphingomyelins, ceramides, triacylglycerols, cholesterol and cholesterol esters) could be analyzed (Figure 6). Significant reduction of sphingomyelins (those containing long aliphatic chains such as 22 and 24 carbon atoms) and significant increase in ceramide content (species N16:0 and N21:0) in plasma of AD patients was observed. From the obtained results it was suggested that the increased ceramide content might result from the accelerated sphingomyelins hydrolysis or increased biosynthesis in brain.

4. Summary. Future directions.
In this review, the great potential of Metabolomics in AD research has been presented. A common
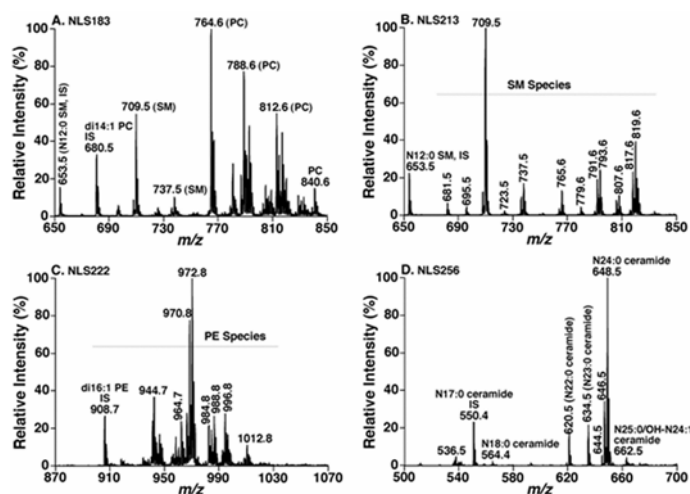
Figure 6. Representative tandem MS analyses of lipid classes present in a human plasma lipid extract (from Han et al., [92]).

conclusion from all metabolomics works is the need to increase the cohort of clinical patients included in practically all the studies. Thus, the metabolomics applications presented in this review are in general considered preliminary works carried out to generate new hypotheses that will need to be corroborated in the future. As a result, sensitivity and specificity of the candidate biomarkers for AD and its progression will need to be evaluated in larger study cohorts to validate these initial results. There is also a great heterogeneity regarding the type of samples (CSF, blood, brain tissue) studied, although, a growing trend on the use of less invasive approach for sample collection has been observed. There is also a clear need of studies investigating biofluid biomarker changes over time in a well-defined (and large enough) group of patients with cognitive deficits. Thus, longer follow-up periods are needed to ascertain whether a patient with stable MCI really does not have incipient AD. It is expected that comprehensive information on clinically relevant metabolites will allow the noninvasive monitoring of AD at pre-dementia or even at "preclinical" stages of AD.

Apart of AD detection from earlier MCI stages, there is a need of the development of robust tools to identify presymptomatic subjects with predisposition and/or presenting risk factors to develop this disease. This detection will help clinicians in selecting those patients who would be most likely to benefit from an early treatment. However, the question is how clinician can select a cohort of patients with presymptomatic neurodegenerative disease. Interesting strategies can be found in literature, for example, the study of preclinical persons at risk for familial Alzheimer's disease can provide a model in which biomarkers can be studied in presymptomatic disease [93]. Although, up to date, presymptomatic stages of AD have not been studied using a metabolomics approaches, future efforts will be needed to identify specifically AD presymptomatic subjects.

Although instrumentation in Metabolomics is expensive, using these approaches hundreds of samples can be analyzed to detect hundreds or thousands of metabolites in a few hours or days. Thus, the costs per sample are relatively low, even more considering the cost of AD diagnosis nowadays. As technology improves, it is most likely that the number of works in this field will continue to increase. As AD has been predominantly studied from a proteomic approach, Metabolomics will potentially contribute in a significant manner to elucidate all the physiopathological processes and ideally may provide therapeutic/preventing mechanisms to brake AD progress.

Certainly, none of the metabolomic works published up to date on the study of AD has integrated omics data from other expression levels (gen, transcript or protein). To understand complex biological systems the integration of the different levels of knowledge will be essential [94, 95]. Information obtained from Metabolomics combined with that obtained from Transcriptomics, Epigenomics and Proteomics would lead to a comprehensive and global molecular picture of the state of a particular biological system at a given time. Integration of "omics" data to study AD is actually demanding the integration of heterogeneous sets of data. To address this data integration, new bioinformatic approaches are under development although integration is nowadays seen as the main bottleneck in the development and application of holistic global approaches that can help us to investigate and understand complex biological systems and diseases including AD [96].

References

[1] Blennow, K., de Leon, M. J., Zetterberg, H., *Lancet* 2006, *368*, 387-403.

[2] Swerdlow, R. H., *Neurotox Res.* 2012, *22*, 182-194.

[3] Povova, J., Ambroz, P., Bar, M., Pavukova, V., Sery, O., Tomaskova, H., Janout, V., *Biomed. Pap.* 2012, *156*, 108-114.

[4] Bateman, R. J., Xong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., Holtzman, D. M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P. S., Ghetti, B., Klunk, W. E., McDade, E., Martins, R. N., Masters, C. L., Mayeux, R., Ringman, J. M., Rossor, M. N., Schofield, P. R., Sperling, R. A., Salloway, S., Morris, J. C., *New Engl. J. Med.* 2012, *367*, 795-80.

[5] Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., Phelps, C. H., *Alzheimers Dement.* 2011, *7*, 270-279.

[6] Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de Leon, M., Feldman, H., Ganguli, M., Hampel H., Scheltens, P., Tierney, M. C., Whitehouse, P., Winblad, B., *Lancet* 2006, *367*, 1262-1270.

[7] Petersen, R. C., *Curr Alzheimer Res.* 2009, *6*, 324-330.

[8] Petersen, R. C., Morris, J. C., *Arch. Neurol.* 2005, *62*, 1160-1163.

[9] Petersen, R. C., *J. Intern. Med.* 2004, *256*, 183-194.

[10] Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R. Jr., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., Wagster, M. V., Phelps, C. H., *Alzheimers Dement.* 2011, *7*, 280-292.

[11] Prince, M., Jackson, J., *World Alzheimer Report 2009*. United Nations Publications, United Kingdom 2009.

[12] Reiman, E. M., Jagust, W. J., *NeuroImage* 2012, *61*, 505-516.

[13] Román, G., Pascual, B., *Archives of Medical Research*. 2012, DOI: 10.1016/j.arcmed.2012.10.018. (In press).

[14] Holtzman, D. M., *J. Mol Neurosci.* 2001, *17*, 147-155.

[15] Blennow, K., Hampel, H., *Lancet Neurol.* 2003, *2*, 605-613.

[16] Craig-Schapiro, R., Fagan, A. M., Holtzman, D. M., *Neurobiol Dis*. 2009, *35*, 128–140.

[17] Flood, D. G., Marek, G. J., Williams, M., *Biochem Pharmacol*. 2011, *81*, 1422-1434.

[18] Hardy, J., Selkoe, D. J., *Science*, 2002, *297*, 353-356.

[19] Blennow, K., Vanmechelen, E., Hampel, H., *Mol Neurobiol.* 2001, *24*, 87-97.

[20] Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V. M., Trojanowski, J. Q., *Ann Neurol.* 2009, *65*, 403-413.

[21] Mattsson, N., Andreasson, U., Persson, S., Arai, H., Batish, S. D., Bernardini, S., Bocchio-Chiavetto, L., Blankenstein, M. A., Carrillo, M. C., Chalbot, S., Coart, E., Chiasserini, D., Cutler, N., Dahlfors, G., Duller, S., Fagan, A. M., Forlenza, O., Frisoni, G. B., Galasko, D., Galimberti, D., Hampel, H., Handberg, A., Heneka, M. T., Herskovits, A. Z., Herukka, S. K., Holtzman, D. M., Humpel, C., Hyman, B. T., Iqbal, K., Jucker, M., Kaeser, S. A., Kaiser, E., Kapaki, E., Kidd, D., Klivenyi, P., Knudsen, C. S., Kummer, M. P., Lui, J., Lladó, A., Lewczuk, P., Li, Q. X., Martins, R., Masters, C., McAuliffe, J., Mercken, M., Moghekar, A., Molinuevo, J. L., Montine, T. J., Nowatzke, W., O'Brien, R., Otto, M., Paraskevas, G. P., Parnetti, L., Petersen, R. C., Prvulovic, D., de Reus, H. P., Rissman, R. A., Scarpini, E., Stefani, A., Soininen, H., Schröder, J., Shaw, L. M., Skinningsrud, A., Skrogstad, B., Spreer, A., Talib, L., Teunissen, C., Trojanowski, J. Q., Tumani, H., Umek, R. M., Van Broeck, B., Vanderstichele, H., Vecsei, L., Verbeek, M. M., Windisch, M., Zhang, J., Zetterberg, H., Blennow, K., *Alzheimers Dement.* 2011, *7*, 386-395.

[22] Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K., Minthon, L., *Lancet Neurol.* 2006, *5*, 228-234.

[23] Aizenstein, H. J., Nebes, R. D., Saxton, J. A., Price, J. C., Mathis, C. A., Tsopelas, N. D., Ziolko, S. K., James, J. A., Snitz, B. E., Houck, P. R., Bi, W., Cohen, A. D., Lopresti, B. J., DeKosky, S. T., Halligan, E. M., Klunk, W. E., *Arch Neurol.* 2008, *65*, 1509-1517.

[24] Cedazo-Minguez, A., Winblad, B., *Exp Gerontol* 2010, *45*, 5-14.

[25] Ghidoni, R., Benussi, L., Paterlini, A., Albertini, V., Binetti, G., Emanuele, E., *Neurodegenerative Dis* 2011, *8*, 413-420.

[26] Dunckley, T., Coon, K. D., Stephan, D. A., *Drug Discov. Today* 2005, *10*, 326-334.

[27] Shi, M., Caudle, W. M., Zhang, J., *Neurobiology Dis*. 2009, *35*, 157-164.

[28] Lista, S., Faltraco, F., Hampel, H., *Progr. Neurobiol.* 2012. DOI: 10.1016/j.pneurobio.2012.06.007. (In Press).

[29] Eckhart, A. D., Beebe, K., Milburn, M., *Clin Transl Sci.* 2012, *5*, 285-288.

[30] Mamas, M., Dunn, W. B., Neyses, L., Goodacre, R., *Arch. Toxicol.* 2011, *85*, 5-17.

[31] Orešič, M., Vidal-Puig, A., Hänninen, V., *Expert Rev. Mol. Diagn.* 2006, *6*, 575-585.

[32] Hassan-Smith, G., Wallace, G. R., Douglas, M. R., Sinclair, A. J., *J. Neuroimmunol.* 2012, *248*, 48-52.

[33] Gowda, G. A. N., Zhang, S., Gu, H., Asiago, V., Shanaiah, N., Raftery, D., *Expert Rev. Mol. Diagn.* 2008, *8*, 617-633.

[34] Zhang, S., Nagana-Gowda, G. A., Ye, T., Raftery, D., *Analyst.* 2010, *135*, 1490-1498.

[35] Lei, Z., Huhman, D. V., Sumner, L. W., *J. Biol. Chem.* 2011, *286*, 25435-25442.

[36] Dettmer, K., Aronov, P. A., Hammock, B. D., *Mass Spectrom. Rev.* 2007, *26*, 51-78.

[37] Rhee, E. P., Gerszten, R. E., *Clin. Chem.* 2012, *58*, 139-147.

[38] O'Connell, T. M., *Bioanalysis* 2012, *4*, 431-451.

[39] Mishur, R. J., Rea, S. L., *Mass Spectrom. Rev.* 2012, *31*, 70-95.

[40] Mandal, R., Guo, A. C., Chaudhary, K. K., Liu, P., Yallou, F. S., Dong, E., Aziat, F., Wishart, D. S., *Genome Med.* 2012, *4,* 38-48.

[41] Nicoli, F., VionDury, J., ConfortGouny, S., Maillet, S., Gastaut, J. L., Cozzone, P. J., *C R Acad Sci III.* 1996, *319*, 623-631.

[42] Ghauri, F. Y. K., Nicholson, J. K., Sweatman, B. C., Wood, J., Beddell, C. R., Lindon, J. C., Cairns, N. J., *NMR Biomed*. 1993, *6*, 163-167.

[43] Kaddurah-Daouk, R., Rozen, S., Matson, W., Han, X., Hulette, C. M., Burke, J. R., Doraiswamy, P. M., Welsh-Bohmer, K. A., *Alzheimers Dement.* 2011, *7*, 309-317.

[44] Samakashvili, S., Ibáñez, C., Simó, C., Gil-Bea, F. J., Winblad, B., Cedazo-Mínguez, A., Cifuentes, A., *Electrophoresis* 2011, *32*, 2757-2764.

[45] Myint, K. T., Aoshima, K., Tanaka, S., Nakamura, T., Oda, Y., *Anal. Chem.* 2009, *81*, 1121-1129.

[46] Czech, C., Berndt, P., Busch, K., Schmitz, O., Wieme,r J., Most, V., Hampel, H., Kastler, J., Senn, H., *PLoS One* 2012, *7*, e31501.

[47] Ibáñez, C., Simó, C., Martín-Álvarez P. J., Kivipelto, M., Winblad, B., Cedazo-Mínguez, A., Cifuentes, A., *Anal. Chem.* 2012, *84*, 8532-8540.

[48] Zhang, A., Sun, H., Wu, X., Wang, X., *Clin Chim Acta.* 2012, *414*, 65-69.

[49] 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., *Nat Protoc.* 2011, *6*, 1060-1083.

[50] Thambisetty, M., Lovestone, S., *Biomark Med.* 2010, *4*, 65-79.

[51] Fonteh, A. N., Harrington, R. J., Tsai, A., Liao, P., Harrington, M. G., *Amino Acids* 2007, *32*, 213-224.

[52] Greenberg, N., Grassano, A., Thambisetty, M., Lovestone, S., Legido-Quigley, C., *Electrophoresis* 2009, *30*, 1235-1239.

[53] Assfalg, M., Bertini, I., Colangiuli, D., Luchinat, C., Schäfer, H., Schütz, B., Spraul, M., *Proc. Natl. Acad. Sci. USA* 2008, *105*, 1420-1424.

[54] Orešič, M., Hyötyläinen, T., Herukka, S. K., Sysi-Aho, M., Mattila, I., Seppänan-Laakso, T., Julkunen, V., Gopalacharyulu, P. V., Hallikainen, M., Koikkalainen, J., Kivipelto, M., Helisalmi, S., Lötjönen, J., Soininen, H., *Transl Psychiatry.* 2011, *1*, e57.

[55] Li N. J., Liu, W. T., Li, W., Li, S. Q., Chen, X. H., Bi, K. S., He, P., *Clin. Biochem.* 2010, *43*, 992-997.

[56] Gonzalez-Dominguez, R., Garcia-Barrera, T., Gomez-Ariza, J. L., *Chem. Pap.* 2012, *66*, 829-835.

[57] Pimplikar, S. W., Nixon, R. A., Robakis, N. K., Shen, J., Tsai, L. H., *J. Neurosci.* 2010, *30*, 14946-14954.

[58] van Echten-Deckert, G., Walter, J., *Progr. Lipid Res.* 2012, *51*, 378-393.

[59] Di Paolo, G., Kim, T. W., *Nat. Rev. Neurosci.* 2011, *12*, 284-296.

[60] Rojo, L., Sjöberg, M. K., Hernández, P., Zambrano, C., Maccioni, R. B., *J. Biomed. Biotechnol.* 2006, *2006*, 73976.

[61] Florent-Béchard, S., Desbène, C., Garcia, P., Allouche, A., Youssef, I., Escanyé, M. C., Koziel, V., Hanse, M., Malaplate-Armand, C., Stenger, C., Kriem, B., Yen-Potin, F. T., Olivier, J. L., Pillot, T., Oster, T., *Biochimie.* 2009, *91*, 804-809.

[62] Sato, Y., Suzuki, I., Nakamura, T., Bernier, F., Aoshima, K., Oda, Y., *J. Lipid Res.* 2012, *53*, 567-576.

[63] Gama-Sosa, M. A., De Gasperi, R., Elder, G. A., *Hum. Genet.* 2012, *131*, 535-563.

[64] Braidy, N., Muñoz, P., Palacios, A. G., Castellano-Gonzalez, G., Inestrosa, N. C., Chung, R. S., Sachdev, P., Guillemin, G. J., *J. Neural. Transm.* 2012, *119*, 173-195.

[65] Trushina, E., Nemutlu, E., Zhang, S., Christensen, T., Camp, J., Mesa, J., Siddiqui, A., Tamura, Y.,Sesaki, H., Wengenack, T. M., Dzeja, P. P., Poduslo, J. F., *PloS one* 2012, *7*, e32737.

[66] Morris, M. C., *Eur. J. Neurol.* 2009, *16*, 1-7.

[67] Wurtman, R. J., Cansev, M., Sakamoto, T., Ulus, I., *Nutr. Rev.* 2010, *68*, S88-S101.

[68] Scheltens, P., Kamphuis, P. J., Verhey, F. R., Olde-Rikkert, M. G., Wurtman, R. J., Wilkinson, D., Twisk, J. W., Kurz, A., *Alzheimers Dement.* 2010, *6*, 1-10.

[69] Peterson, J. J., Dwyer, J. T., Jacques, P. F., McCullough, M. L., *Nutr. Rev.* 2012, *70*, 491-508.

[70] Finks, S. W., Airee, A., Chow, S. L., Macaulay, T. E., Moranville, M. P., Rogers, K. C., Trujillo, T. C.,

Pharmacotherapy 2012, *32*, 54-87.

[71] Choi, D. Y., Lee, Y. J., Hong, J. T., Lee, H. J., *Brain Res. Bull.* 2012, *87*, 144-153.

[72] Mancuso, C., Siciliano, R., Barone, E., Preziosi, P., *Biochim Biophys Acta* 2012, *1822*, 616-624.

[73] Ebrahimi, A., Schluesener, H., *Ageing Res. Rev.* 2012, *11*, 329-345.

[74] Wang, J., Ferruzzi, M. G., Ho, L., Blount, J., Janle, E. M., Gong, B., Pan, Y., Gowda, G. A. N., Raftery, D., Arrieta-Cruz, I., Sharma, V., Cooper, B., Lobo, J., Simon, J. E., Zhang, C., Cheng, A., Qian, X., Ono, K., Teplow, D. B., Pavlides, C., Dixon, R. A., Pasinetti, G. M., *Neurosci.* 2012, *32*, 5144-5150.

[75] Jung, H. R., Sylvänne, T., Koistinen, K. M., Tarasov, K., Kauhanen, D., Ekroos, K., *Biochim Biophys Acta* 2011, *1811*, 925-934.

[76] Orešič, M., *Biochim Biophys Acta* 2011, *811*, 991-999.

[77] Han, X., Yang, K., Gross, R. W., *Mass Spectrom Rev.* 2012, *31*, 134-178.

[78] Griffin, J. L., Nicholls, A. W., *Pharmacogenomics* 2006, *7*, 1095-1107.

[79] Hua, C., van der Heijden, R., Wang, M., van der Greef, J., Hankemeier, T., Xu, G., *J. Chromatogr B* 2009, *877*, 2836-2846.

[80] Wenk, M. R., *Cell* 2010, *143*, 888-895.

[81] Raber, J., Huang, Y., Ashford, J. W., *Neurobiol. Aging* 2004, *24*, 641-650.

[82] Cedazo-Minguez, A., Cowburn, R. F., *J. Cell. Mol. Med.* 2001, *5*, 254-266.

[83] Cedazo-Mínguez, A., *J. Cell. Mol. Med.* 2007, *11*, 1227-1238.

[84] Han, X., *Curr. Alzheimer Res.* 2005, *2*, 65-77.

[85] Han, X., *BBA-Lipid Lipid Met.* 2010, *1801*, 774-783.

[86] Blanksby, S. J., Mitchell, T. W., *Annu. Rev. Anal. Chem.* 2010, *3*, 433-465.

[87] Nygren, H., Seppänen-Laakso, T., Castillo, S., Hyötyläinen, T., Orešič, M., *Methods Mol. Bio.* 2011, *708*, 247-257.

[88] Li, M., Zhou, Z., Nie, H., Bai, Y., Liu, H., *Anal Bioanal Chem.* 2011, *399*, 243-249.

[89] Han, X., Holtzman, D. M., McKeel D. W. Jr., *J. Neurochem.* 2001, *77*, 1168-1180.

[90] Han, X., Holtzman, D. M., McKeel D. W. Jr., Kelley, J., Morris, J. C., *J. Neurochem.* 2002, *82*, 809-818.

[91] Han, X., *Neurochem.* 2007, *103*, 171-179.

[92] Han, X., Rozen, S., Boyle, S. H., Hellegers, C., Cheng, H., Burke, J. R., Welsh-Bohmer, K. A., Doraiswamy, P. M., Kaddurah-Daouk, R., *PLoS ONE* 2011, *6*, 1-13.

[93] Ringman, J. M., Younkin, S. G., Pratico, D., Seltzer, W., Cole, G. M., Geschwind, D. H., Rodriguez-Agudelo, Y., Schaffer, B., Fein, J., Sokolow, S., Rosario, E. R., Gylys, K. H., Varpetian, A., Medina, L. D., Cummings, J. L., *Neurology* 2008, *71*, 85-92.

[94] Ibáñez, C., Valdés, A., García-Cañas, V., Simó, C., Celebier, M., Rocamora-Reverte, L., Gómez-

Martínez, A., Herrero, M., Castro-Puyana, M., Segura-Carretero, A., Ibáñez, E., Ferragut, J. A., Cifuentes, A*, J Chromatogr A* 2012, *1248*, 139-153.

[95] García-Cañas, V., Simó, C., Herrero, M., Ibáñez, E., Cifuentes, A., *Anal Chem.* 2012, *84*, 10150-10159.

[96] Orešič, M., Lötjönen, J., Soininen, H., *Genome Med.* 2010, *2*, 83-87.

4.3.   CAPÍTULO 5 (Chapter 5)

Analysis of chiral amino acids in cerebrospinal fluid samples linked to different stages of Alzheimer's disease

2757

Shorena Samakashvili[1]
Clara Ibáñez[2]
Carolina Simó[2]
Francisco J. Gil-Bea[3]
Bengt Winblad[3]
Angel Cedazo-Mínguez[3]
Alejandro Cifuentes[2]

[1]Department of Physical and
 Analytical Chemistry, School of
 Exact and Natural Sciences,
 Tbilisi State University, Tbilisi,
 Georgia
[2]Laboratory of Foodomics,
 Institute of Food Science
 Research (CSIC), Nicolas
 Cabrera, Madrid, Spain
[3]NVS Department, KI-Alzheimer's
 Disease Research Center,
 Karolinska Institute, Stockholm,
 Sweden

Received March 1, 2011
Revised April 13, 2011
Accepted April 15, 2011

Research Article

Analysis of chiral amino acids in cerebrospinal fluid samples linked to different stages of Alzheimer disease

Chiral micellar electrokinetic chromatography with laser-induced fluorescence detection (chiral-MEKC-LIF) was used to investigate D- and L-amino acid contents in cerebrospinal fluid (CSF) samples related to different Alzheimer disease (AD) stages. CSF samples were taken from (i) control subjects (S1 pool), (ii) subjects showing a mild cognitive impairment who remained stable (S2 pool), (iii) subjects showing an mild cognitive impairment that progressed to AD (S3 pool) and (iv) subjects diagnosed with AD (S4 pool). The optimized procedure only needed 10 μL of CSF and it included sample cleaning, derivatization with FITC and chiral-MEKC-LIF separation. Eighteen standard amino acids were baseline separated with efficiencies up to 703 000 plates/m, high sensitivity (LODs in the nM range) and good resolution (values ranging from 2.6 to 9.5). Using this method, L-Arg, L-Leu, L-Gln, γ-aminobutyric acid, L-Ser, D-Ser, L-Ala, Gly, L-Lys, L-Glu and L-Asp were detected in all the CSF samples. S3 and S4 samples (i.e. AD subjects) showed significant lower amounts of L-Arg L-Lys, L-Glu and L-Asp compared to the non-AD S1 and S2 samples, showing in the S4 group the lowest amounts of L-Arg L-Lys, L-Glu and L-Asp. Moreover, γ-aminobutyric acid was significantly higher in AD subjects with the highest amount also found for S4. No significant differences were observed for the rest of amino acids including D-Ser. Based on the obtained chiral-MEKC-LIF data, it was possible to correctly classify all the samples into the four groups. These results demonstrate that the use of enantioselective procedures as the one developed in this work can provide some new light on the investigations of AD, including the discovery of new biomarkers related to different stages of AD.

Keywords:

Alzheimer / CE / Cerebrospinal fluid / Chiral amino acids

DOI 10.1002/elps.201100139

1 Introduction

Amino acids have been the focus of much attention in biomedical research, medical diagnostics, clinical chemistry and the pharmaceutical industry, because they play essential roles in control and regulation of crucial functions in the human body [1, 2]. Moreover, it has recently been found that L- and D-amino acids can play different biological functions and, therefore, analysis of these chiral forms in biological fluids may provide a means of diagnosis and possible

Correspondence: Professor Alejandro Cifuentes, Laboratory of Foodomics, Institute of Food Science Research (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain
E-mail: a.cifuentes@csic.es
Fax: +34-91-5644853

Abbreviations: AD, Alzheimer disease; **CNS,** central nervous system; **CSF,** cerebrospinal fluid; **FSDA,** forward stepwise discriminant analysis; **GABA,** γ-aminobutyric acid; **MCI,** mild cognitive impairment

treatment of diseases [3–5]. However, the knowledge of D-amino acids in biological samples is still limited and most of the D-amino acids are not well investigated yet, concluding that more work is needed in this area of research [5].

Alzheimer disease (AD) is a neurodegenerative disorder with an estimated worldwide prevalence of over 18 million people, and is predicted to increase with an increasing elderly population [6]. AD is characterized by cognitive deficits and memory impairment, and there is currently no cure for this disease. Biochemical processes accounting for neurodegeneration are not known but are likely to include the metabolism of amino acids [6]. Since free amino acids are important in neurotransmission, receptor function and are implicated in neurotoxicity, changes in free amino acids metabolism can be an early indicator of neurodegeneration in AD. This early diagnose is considered crucial for a potential AD cure because treatment might be most effective when initiated very early in the course of AD, before amyloid plaques and neurodegeneration become too widespread. Thus, biomarkers are needed that can detect

www.electrophoresis-journal.com

AD in the predemencia phase or, ideally, in presymptomatic individuals [7, 8]. Moreover, in vitro assays have recently shown that aggregation and folding parameters of amyloid beta are stereospecific and the aggregation property strongly depends upon the amino acid sequence and their stereo-specificity [9], suggesting the stereospecific role of amino acids comprising aggregation and its relevance to neurodegeneration.

Analysis of chiral amino acids can, therefore, be an interesting strategy to further investigate AD. Moreover, analysis of the chiral forms of amino acids can provide more light on the contradictory results usually found in the literature on amino acids analysis related to AD. For instance, some researchers have found no differences in glutamate levels in the hippocampus and cerebral cortices between normal and AD patients [10], whereas other researchers have shown that this amino acid was elevated [11] or reduced [12, 13] in brain of AD patients. Similarly, aspartate levels have been found increased [14], reduced [12, 13] or virtually unaltered in AD compared to normal brain [11]. The same differing results are found comparing the published works dealing with the analysis of amino acids in cerebrospinal fluid (CSF) from AD and control subjects using chromatographic [6, 15–19] or electrodriven separation techniques [6, 20–23]. Though further studies are needed to clarify whether the alterations in amino acids in an AD brain and CSF is a factor or a result, the elucidation of the relationship of amino acids (including their L- and D-forms) and AD might offer new insights into overcome this devastating disease.

In this regard, chiral CE can be a good choice to carry out the analysis of D- and L-amino acids from CSF samples based on CE high separation efficiency, enormous resolving power, large peak capacities, short analysis times and small sample and reagents volume (including tiny amounts of chiral selectors) [24]. Although, as mentioned above, some previous works have already shown the content of amino acids in CSF using different techniques such as HPLC [6, 15–17] or CE [6, 20, 22], to our knowledge, only two papers have faced the complex issue regarding the analysis of chiral amino acids in CSF samples related to AD. Fisher et al. showed in a first work using HPLC that free D-Asp was significantly higher ($p < 0.01$) in AD ventricular CSF compared to normal ventricular CSF [25]. However, in a posterior work, they found that also D-Ser was significantly higher ($p < 0.01$) in AD ventricular CSF [26]. They concluded that CSF could reflect the degenerative process that occurs in AD since CSF is the repository of amino acids from the brain, although, they state that more work is needed to corroborate this crucial point [26].

The goal of this work is, therefore, to carry out the profiling of the main D- and L-amino acids that can be found in CSF samples related to different stages of AD. To do this, a new analytical CE method is developed combining micellar electrokinetic chromatography (MEKC) with a chiral selector and laser induced fluorescence (LIF) detection to analyze a group of 18 selected L- and D-amino acids

that can be found in CSF. The developed chiral-MEKC-LIF method is fast and reproducible and allows the separation of the 18 amino acids with good efficiency and sensitivity.

2 Materials and methods

2.1 Chemicals

All chemicals were of analytical reagent grade and used as received. β-CD was used as chiral selector for the MEKC running buffer together with SDS and boric acid, all from Sigma-Aldrich (St. Louis, MO, USA). Water was deionized by using a Milli-Q system from Millipore (Bedford, MA, USA). An aqueous solution containing 5 mol/L of sodium hydroxide from Panreac Quimica S.A. (Barcelona, Spain) was used to adjust the pH of the BGEs. A 0.1 mol/L NaOH solution was used to rinse the capillary. BGEs were stored at 4°C and warmed at room temperature before use. All solutions were filtrated through a 0.45 μm membrane filter. A 20 mM FITC solution from Sigma-Aldrich was prepared in acetone from Merck (Darmstadt, Germany). Standard L- and D-amino acids were from Sigma-Aldrich. Stock solutions of 0.05 mM Gly, 0.10 mM D/L-Arg, D/L-Ser, D/L-Leu, D/L-Ala, D/L-Gln, D/L-Glu, γ-aminobutyric acid (GABA) and 0.50 mM DL-Lys and DL-Asp were prepared in Milli-Q water. The stock solutions were stored at 4°C.

2.2 CSF samples

2.2.1 Study population

The patients included in the study ($n = 99$) were from the Memory Clinic at the Karolinska University Hospital in Huddinge (Sweden). Thirty-three had just subjective cognitive impairment and they were considered as control group (sample 1, S1), 39 had mild cognitive impairment (MCI) of which 26 remaining stable (sample 2, S2) and 13 had progression-to-AD within 2 years' time (sample 3, S3), the fourth group was composed of 27 patients already presenting mild AD (sample 4, S4). These patients were all living independently in the community. They were evaluated according to a standard comprehensive assessment protocol including clinical examination, brain imaging, electroencephalography, analyses of blood and CSF (including total tau (T-Tau), phospho-tau (P-Tau) and Aβ$_{1-42}$) and a detailed neuropsychological evaluation. Dementia and AD were diagnosed according to Diagnostic and Statistical Manual as Mental Disorders, Fourth Edition and National Institute of Neurological and Communicative Disorders and the Alzheimer's Disease and Related Disorders Association criteria. MCI patients from S2 were not demented, had (self and/or an informant) reported cognitive decline and impairment on objective cognitive tasks, and S3 had preserved basic AD/minimal impairment in complex

instrumental functions. S1 patients had cognitive complaints without impairment on objective cognitive tasks. Patients with psychiatric disorders (i.e. depression, alcohol abuse) or other conditions (i.e. diabetes, brain tumours, normal pressure hydrocephalus) were not included. The study was conducted under the guidelines of the Declaration of Helsinki and approved by the ethics committee of the Karolinska Institutet.

2.2.2 CSF extraction

CSF samples were obtained by lumbar puncture performed in the sitting position. CSF extraction is routinely performed at the Karolinska University Hospital Memory clinic in Huddinge (Sweden) as part of the medical examination. The extractions were performed in the mornings in fasting patients. CSF samples were obtained from L3/L4 or L4/L5 interspaces after local anaesthetic infiltration in the skin. After disposal of the first millilitre the following 10 mL were collected in polypropylene tubes. No sample containing more than 500 erythrocytes/μL CSF samples was used. Samples were gently mixed to avoid gradient effects and centrifuged at $2000 \times g$ for 10 min to eliminate cells and insoluble material. Supernatants were immediately aliquoted, immediately frozen and stored at -80°C for pending biochemical analyses. In this study, pools of each of the four groups were used. Pools were obtained by mixing 10 μL of each individual sample, vortexed and immediately stored at -80°C.

2.2.3 Ultrafiltration

Aliquots of 100 μL of each sample (S1, S2, S3 and S4) was centrifuged at $14000 \times g$ for 25 min on a 3 kDa membrane to remove proteins from CSF samples. The <3 kDa fraction was used for subsequent FITC derivatization procedure.

2.3 Derivatization procedure

The FITC derivatization procedure was optimized as described below. The selected conditions consisted of mixing 10 μL of the standard solution of amino acids or CSF sample with 80 μL of a 25 mM sodium borate buffer at pH 10.0 and 10 μL of a 20 mM FITC solution (freshly prepared just before derivatization). The reaction took place overnight in darkness at room temperature. After derivatization, the solution was stored in dark at 4°C before chiral-MEKC-LIF analysis.

2.4 MEKC-LIF conditions

All analyses were carried out in triplicate using a P/ACE 2100 CE apparatus from Beckman Instruments (Fullerton, CA, USA) equipped with an Ar+ laser at 488 nm (excitation wavelength) and 520 nm (emission wavelength), also from

Beckman Instruments. Bare fused-silica capillary was purchased from Composite Metal Services (Worcester, England). The capillary dimensions were 50 cm of detection length, 57 cm of total length, and 50-μm id and was thermostated at 30°C. Injections were made at the anodic end using N_2 at 0.5 psi (3.45 kPa) for 3 s (hydrodynamic injection of 3.5 nL), and the applied voltage was $+20$ kV. The P/ACE 2100 CE instrument was controlled by a PC running the System GOLD software from Beckman. Before first use, new capillaries were preconditioned by rinsing with 0.1 M NaOH for 30 min. The washing protocol between runs was optimized to obtain adequate repeatability, selecting the following conditions: at the beginning of each run, the capillary was rinsed with 0.1 M NaOH for 2 min, followed by 2 min with Milli-Q water and then equilibrated for 5 min with the running buffer. After optimization, 100 mM sodium tetraborate, 80 mM SDS and 20 mM β-CD at pH 10.0 was used as running buffer. In order to skip any irreproducibility problem derived from buffer depletion, the buffer vial was changed every three injections. At the end of the day, the capillary was rinsed with Milli-Q water for 10 min, and then nitrogen was passed for 2 min.

2.5 Data analysis

Statistical analysis was performed using Statistica 7.1 software (StatSoft 2005, USA, www.statsoft.com). First, one-way ANOVA of the amino acid corrected peak areas (i.e. peak area/migration time) was carried out to detect significant differences among the four groups of samples, using a 1% significance level and a Scheffe test for means comparison. Then, a forward stepwise discriminant analysis (FSDA) of the corrected peak areas was applied to select the variables most useful to differentiate the four groups of samples and to obtain the classification functions. Finally, a Fisher's canonical variable analysis (Root 1 and Root 2) was performed in order to obtain a low-dimensional graphical representation of the samples separating as much as possible the groups under study.

3 Results and discussion

A group of eight chiral amino acids (D/L-Arg, D/L-Ser, D/L-Leu, D/L-Ala, D/L-Gln, D/L-Glu, D/L-Lys and D/L-Asp) plus the nonchiral amino acids Gly and GABA was initially selected to carry out this study. This selection was based on a literature search on the main amino acids usually found in CSF samples.

3.1 Development of the FITC derivatization procedure

Previous to the optimization of the chiral-MEKC-LIF method, the derivatization procedure was studied in order to achieve the maximum amino acid signal with the lowest

number of interferences from the derivatizing reagent FITC, which is known to produce a high number of interfering fluorescent compounds [27]. Besides, different parameters were tested in order to obtain a sensitive and robust derivatizing procedure of the tiny volume of CSF sample used in this work (10 μL). Different concentrations of FITC solution were tested (i.e. to give a final concentration equal to five, ten and 20 times greater than the concentration of amino acids); different pHs (9–11) and concentrations of sodium tetraborate buffer (from 20–400 mM) were tested. Optimum results in terms of sensitivity and reproducibility of the derivatization method of the 10 μL sample were obtained using a FITC solution with a concentration ten times greater than the solution of amino acids, and in a 25 mM sodium tetraborate buffer at pH 10.0.

3.2 Optimization of chiral-MEKC-LIF separation conditions and figures of merit

Chiral-MEKC-LIF method was optimized using β-CD as chiral selector at different concentrations (10, 15 and 20 mM) and testing different SDS concentrations (40, 60, 80 and 100 mM), pH buffer values (9.5, 9.7, 10.0, 10.3) and sodium tetraborate concentrations (25, 50, 100, 120 mM). Optimum chiral resolution was obtained for the 18 amino acids investigated in this work using 100 mM sodium tetraborate, 80 mM SDS and 20 mM β-CD solution at pH 10.0 as BGE.

A typical electropherogram of the amino acids mixture obtained under these conditions is shown in Fig. 1. As can be seen, using this BGE it is possible to achieve the complete chiral-MEKC-LIF separation of D- and L-Arg, D- and L-Leu, D- and L-Gln, GABA, D- and L-Ser, D- and L-Ala, Gly, D- and L-Glu, D- and L-Lys, as well as D- and L-Asp with resolutions higher than 2.6 in all the cases (see Table 1). Thus, these new analytical conditions bring about the baseline separation for the 18 amino acids as clearly shown in Fig. 1. Besides, they are also well separated from the impurities coming from the FITC derivatization (peaks marked with an asterisk). Moreover, efficiencies ranging from 242 000 plates/m for L-Asp to 703 000 plates/m for D-Arg were achieved corroborating the usefulness of this procedure. Also, the LODs, calculated considering a signal to noise ratio equal to 3, ranged from 0.8 nM for Gly to 16.5 nM for L-Asp, providing LOQ ranging from 2.6 nM for Gly to 53.6 nM for L-Asp. The method was determined to be reproducible according to the results given in Table 2. Thus, %RSD values obtained for intra-day repeatability (five consecutive injections in the same day) were better than 5.3 and 0.5% for peak areas and migration time, respectively. On the other hand, RSD values obtained for three different days ($n = 15$) were better than 6.6 and 2.9% for peak areas and migration time, respectively, assuring an adequate repeatability of the analysis.

In spite of these good results, the present method can sill be improved as can be deduced from the numerous
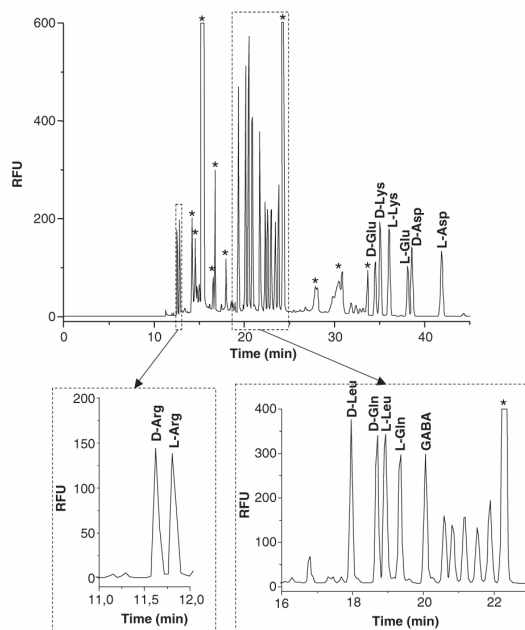
Figure 1. MEKC-LIF electropherograms of a standard mixture of 18 D/L amino acids (10 μg/mL D/L-Arg, D/L-Leu, D/L-Gln, GABA, D/L-Ser, D/L-Ala, D/L-Glu, 5 μg/mL Gly, 50 μg/mL D/L-Lys and D/L-Asp). Peaks marked with an asterisk correspond to FITC peaks. BGE: 100 mM sodium tetraborate, 80 mM SDS, 20 mM β-CD at pH 10.0. FITC-derivatized amino acids injected for 3 s at 0.5 psi; bare fused-silica capillary with 57 cm total length, 50 cm detection length and 50 μm id; running voltage, 20 kV; LIF detection, Ar+ laser at 488 nm (excitation wavelength) and 520 nm (emission wavelength).

impurities that are obtained from the FITC derivatization (see peaks marked with an asterisk in Fig. 1) that can interfere in the determination of other amino acids not included in the standard mixture. However, given the good figures of merit in terms of resolution, efficiency and sensitivity of this chiral-MEKC-LIF procedure, this method was used to carry out the chiral amino acids profiling in CSF samples.

3.3 Identification and relative quantitation of D/L-amino acids in CSF samples related to different AD stages

Amino acids from CSF samples related to different AD stages were derivatized and analyzed by chiral-MEKC-LIF using the selected conditions described above. The investigated CSF samples were taken from control subjects (S1 pool, $n = 33$ individuals), subjects showing an MCI who remained stable (S2 pool, $n = 26$ individuals), subjects showing an MCI that progressed to AD (S3 pool, $n = 13$ individuals) and subjects diagnosed with AD (S4 pool, $n = 27$ individuals). In order to identify the different

Table 1. Figures of merit of the chiral-MEKC-LIF method developed in this work
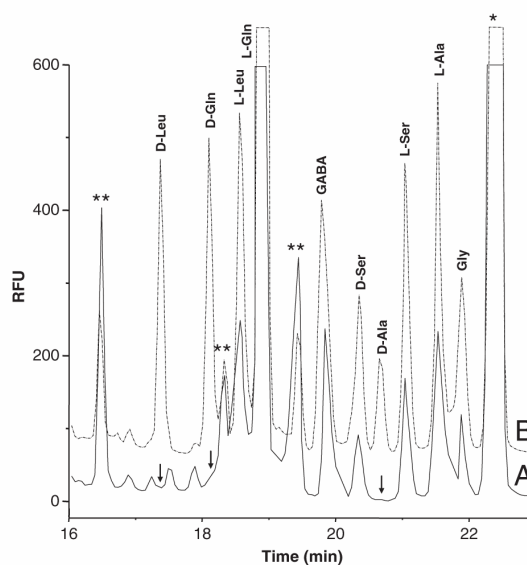
Amino acid	LOD (nM)[a]	Efficiency[b]	Resolution[c]
D-Arg	1.6	703 000	2.6
L-Arg	1.6	686 000	
D-Leu	0.9	492 000	7.0
L-Leu	0.7	467 000	
D-Gln	0.9	433 000	4.4
L-Gln	1.0	440 000	
GABA	1.1	419 000	–
D-Ser	1.9	441 000	3.5
L-Ser	1.9	444 000	
D-Ala	2.0	363 000	4.0
L-Ala	2.3	417 000	
Gly	0.8	384 000	–
D-Glu	3.4	299 000	9.5
L-Glu	4.4	282 000	
D-Lys	11.7	292 000	2.7
L-Lys	12.3	297 000	
D-Asp	15.4	272 000	7.3
L-Asp	16.5	242 000	

a) LOD calculated for a signal/noise ratio of 3.
b) Efficiency (*N*) calculated as number of theoretical plates per meter of column.
c) Resolution (*R*) between two peaks (namely, 1 and 2) was calculated using the next formula: $R = (t_2 - t_1)/(w_2 - w_1)$, where t_1 and t_2 are the analysis time of enantiomers 1 and 2, respectively, and w_1 and w_2 are the widths at the base of enantiomers 1 and 2, respectively.

Table 2. Time analysis and corrected peak area reproducibility (determined as RSD, %RSD) of the chiral-MEKC-LIF method for the same day (*n* = 5) and three different days (*n* = 15)

Amino acid	%RSD (intraday)		%RSD (3 days)	
	t_{mig}	Corrected peak area	t_{mig}	Corrected peak area
D-Arg	0.3	3.8	2.0	5.0
L-Arg	0.3	4.2	2.0	5.0
D-Gln	0.3	3.0	2.1	5.2
L-Gln	0.5	3.9	2.1	5.8
D-Ser	0.5	2.8	2.3	5.9
L-Ser	0.4	4.5	2.4	6.1
D-Asp	0.5	4.8	2.7	6.3
L-Asp	0.3	5.3	2.9	6.6

Figure 2. MEKC-LIF analysis of CSF sample S2 (solid trace A) and CSF sample spiked with D/L-Leu, D/L-Gln, GABA, D/L-Ser, D/L-Ala and Gly (dotted trace B). Peaks marked with *correspond to FITC-related peaks and those marked with **correspond to unidentified peaks from CSF sample. All the analytical conditions were as in Fig. 1.

enantiomers, a co-injection procedure as shown in Fig. 2 was applied allowing the accurate identification of 11 free L/D-amino acids in all the CSF samples. L-Arg, L-Leu, L-Gln, GABA, L-Ser, D-Ser, L-Ala, Gly, L-Lys, L-Glu and L-Asp were identified in all the investigated CSF samples. As can be deduced from our results, the only D-amino acid found in CSF was D-Ser. Interestingly, D-Ser is thought to have an important function in the central nervous system (CNS) of mammals to modulate the *N*-methyl-D-aspartate subtype of

glutamate receptor [5, 28, 29]. Thus, the finding of D-Ser is in good agreement with the fact that only for this D-amino acid it has been possible to demonstrate its natural origin based on the activity of the enzyme serine racemase that catalyzes the direct racemization of L-Ser to D-Ser [5]. So far, there is not a clear explanation for other D-amino acids that have sometimes been found in mammalians including CSF samples, suggesting that they can be derived from nutrition or in many biological samples from bacteria [5].

As already mentioned, further studies are needed to clarify whether the alterations in D-amino acids in AD CSF is a factor or a result. In this regard, a correct clinical diagnosis of AD early in the course of the disease is crucial to initiate symptomatic treatment, and will be even more important when disease-arresting drugs, such as β-sheet breakers or γ-secretase inhibitors, will reach the clinic. However, there is no clinical method to determine if a patient with MCI has incipient AD, i.e. will progress to AD with dementia, or have a benign form of MCI without progression [30]. Thus, there is a great clinical need for diagnostic biomarkers to identify incipient AD in MCI cases. With this idea in mind, our next step was to inject the CSF samples from the four different groups studied in this work, which included control subjects (S1), subjects showing MCI who remained stable (S2), subjects showing MCI that progressed to AD (S3) and subjects already diagnosed with AD (S4).

The relative levels of the 11 amino acids found in the CSF samples from S1 to S4 groups are given in Table 3, together with their standard deviation after triplicate

analysis. Mean values were also compared and the results are included as superscripts. Some interesting differences among the CSF samples are observed in Table 3. Thus, using the statistical approach described in Data analysis (vide supra) the most significant variables ($p < 0.01$) able to differentiate S1-S2 (non-AD subjects) from S3-S4 (AD subjects) were confirmed to be L-Glu, L-Arg, L-Asp, L-Lys and GABA. Thus, S3 and S4 samples (i.e. AD subjects) showed significant ($p < 0.01$) lower amounts of L-Arg, L-Lys, L-Glu and L-Asp compared to the non-AD samples S1 and S2, observing in the S4 sample the lowest amounts of L-Arg L-Lys and L-Glu. Similar results were mentioned by D'Aniello et al. [17] who observed that aspartic acid occurred at significantly lower concentrations in CSF from AD patients than in CSF from healthy individuals. Also, these results seem to be in good agreement with the excitatory role of some of these amino acids. Thus, as the major excitatory neurotransmitters in the mammalian CNS, Glu and Asp are present in more than half of all CNS synapses, which underscores their important involvement in, e.g. learning, memory or movement [31, 32]. Based on the observed decrease of these crucial functions in AD patients, lower levels of L-Glu and L-Asp in AD patients could be expected as corroborated by results in Table 3.

Moreover, GABA was significantly higher ($p < 0.01$) in AD subjects from S3 and S4 compared to S1 and S2, with the highest amount of GABA also found in S4. Our results do not agree with those reported by Bergquist et al. [20], who analyzed CSF samples from patients with AD and other neurological diseases, observing that GABA levels were reduced in the CSF of AD patients, while Gly levels in the AD patients were elevated. This disagreement can be explained considering that GABA and Gly are the main inhibitory neurotransmitters in CNS [33]. In fact, as many as 10–40% of nerve terminals in the hippocampus and cerebral cortex may use GABA as a neurotransmitter to transmit "closure" signals [22, 34]. Therefore, GABA and/or Gly can be expected to increase in AD patients considering their main activity as inhibitory neurotransmitters of the CNS. This finding is also supported by Lanctot et al. [35] who hypothesized that the variable findings regarding GABA disruption in AD patients would reflect subtypes of this disease that could possibly be manifested clinically by differing behavioural symptoms.

No significant differences ($p < 0.01$) were observed for the rest of amino acids including D-Ser. Interestingly, the result on D-Ser does not agree with those reported by Fisher et al. [25, 26] who in the 90's found using HPLC that free D-Ser in CSF from AD patients was significantly higher than in normal subjects. Since then, to our knowledge, the present work is the only published study on the level of D- and L-amino acids in CSF samples related to AD and it seems to rebut the accepted hypothesis of a higher level of D-Ser in AD CSF. Moreover, similar results were found by other authors in 1993 [36] and 1995 [37] using HPLC; they did not observe significant differences in the level of D-Ser in

Table 3. Mean values ($\pm$SD) of relative levels of L- and D-amino acids found in the four different CSF pools analyzed

Amino acid	CSF-S1 pool	CSF-S2 pool	CSF-S3 pool	CSF-S4 pool
	Rel. level[a] ($\pm$SD)[b]	Rel. level ($\pm$SD)	Rel. level ($\pm$SD)	Rel. level ($\pm$SD)
D-Arg	ND[c]	ND	ND	ND
L-Arg	1423[b] $\pm$ 113	1243[b] $\pm$ 63	868[a] $\pm$ 44	802[a] $\pm$ 16
D-Leu	ND	ND	ND	ND
L-Leu	839[a] $\pm$ 51	1107[b] $\pm$ 108	1253[b] $\pm$ 40	1295[b] $\pm$ 40
D-Gln	ND	ND	ND	ND
L-Gln	9174[a] $\pm$ 449	9373[a] $\pm$ 101	8933[a] $\pm$ 166	9161[a] $\pm$ 120
GABA	1137[a] $\pm$ 108	1370[ab] $\pm$ 22	1611[bc] $\pm$ 22	1805[c] $\pm$ 162
D-Ser	274[a] $\pm$ 17	266[a] $\pm$ 23	236[a] $\pm$ 21	239[a] $\pm$ 13
L-Ser	779[a] $\pm$ 30	769[a] $\pm$ 22	731[a] $\pm$ 36	715[a] $\pm$ 25
D-Ala	ND	ND	ND	ND
L-Ala	1295[a] $\pm$ 134	1302[a] $\pm$ 57	1258[a] $\pm$ 114	1189[a] $\pm$ 47
Gly	588[b] $\pm$ 51	328[a] $\pm$ 12	332[a] $\pm$ 15	298[a] $\pm$ 29
D-Lys	ND	ND	ND	ND
L-Lys	173[c] $\pm$ 5	160[bc] $\pm$ 11	139[b] $\pm$ 10	104[a] $\pm$ 7
D-Glu	ND	ND	ND	ND
L-Glu	87[b] $\pm$ 4	84[b] $\pm$ 2	43[a] $\pm$ 6	39[a] $\pm$ 2
D-Asp	ND	ND	ND	ND
L-Asp	29[b] $\pm$ 2	30[b] $\pm$ 2	17[a] $\pm$ 1	19[a] $\pm$ 1

a) Relative level given in arbitrary units and calculated as corrected peak area (peak area/migration time).
b) SD ($n = 3$).
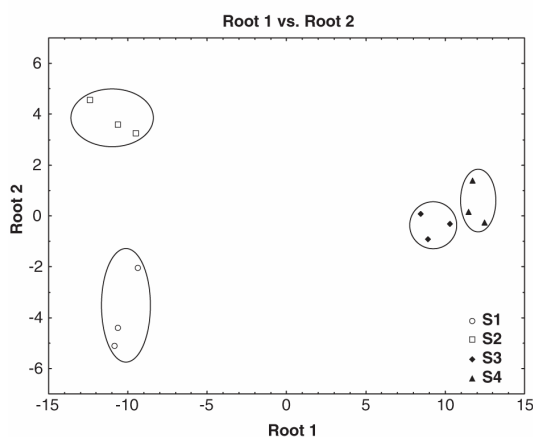c) Not detected.
Rows without a common superscript letter are significantly different ($p < 0.01$). CSF samples were taken from control subjects (S1 pool), subjects showing an MCI who remained stable (S2 pool), subjects showing an MCI that progressed to AD (S3 pool) and subjects diagnosed with AD (S4 pool).

normal and AD human brain. In this regard, some authors have mentioned that CSF contamination with erythrocytes could also increase D-Ser concentration [38]. In our opinion, when combining the results from the above studies, there is no consistent D-Ser disruption in AD CSF samples. Thus, D-Ser should not be considered at this moment a reliable diagnostic marker for AD when analyzing CSF samples.

Although a larger number of samples have to be analyzed in order to confirm the above findings, it is interesting to remark that significant difference in all the investigated samples was only observed for five compounds out of the 18 investigated. Thus, other two compounds differing more than statistically expected were Gly and L-Leu. However, the statistically significant difference for Gly and L-Leu is not systematically observed for all the groups as can be deduced from the results given in Table 3.

To confirm the conclusions obtained from the statistical analysis discussed above, all the data as a whole were subjected to an FSDA. The data matrix has dimensionality 12 · 7 (seven chiral-MEKC-LIF peak areas chosen from electropherograms of the four CSF groups, S1–S4, run in triplicate), data analysis was carried out as described in the Section 2. In Fig. 3, each point represents one particular chiral-MEKC-LIF electropherogram confirming the straightforward statistical analysis performed above. L-Glu, L-Arg, L-Asp, L-Lys and GABA were selected as the most important variables to differentiate the four groups of samples. Values of 4.0 and 3.9 were considered for *F*-statistic to enter and to remove variables, respectively. With these five amino acids, it was possible to classify all samples correctly. As can be seen in Fig. 3, the four groups S1–S4 are nicely separated using this approach. The classification

matrix of this analysis reveals L-Glu, L-Asp and L-Lys as crucial for the four groups' sorting. The factor structure matrix associated to the FSDA highlight L-Glu and L-Arg as the amino acids that most contribute to AD and non-AD classification. S1 and S2 (non AD samples) can be clearly separated from S3 and S4 (AD samples) only by analyzing the levels of these two compounds. Interestingly, the decrease observed in L-Glu and L-Arg in AD CSF samples seems to be in good agreement with the much higher amount of GABA found in the same samples, since GABA is produced from L-Arg via their conversion to ornithine and L-Glu by the action of the enzyme arginase also found in the brain [39–41]. However, some other possible explanations of the decrease of L-Arg in AD CSF samples cannot be ruled out, including a major concentration or activity of enzymes using L-Arg as substrate as, e.g. nitric oxide synthase or peptidylarginine deiminases [41, 42].

These results seem to demonstrate the good possibilities of chiral-MEKC-LIF analysis of CSF samples as a clinical diagnostic tool of AD even at the early stages of this disease. Moreover, as can be deduced from the good separation between groups S2 and S3 it seems that chiral-MEKC-LIF analysis of CSF samples could be proposed as a clinical method to determine if a patient with MCI has incipient AD, i.e. will progress to AD with dementia (group S3) or have a benign form of MCI without progression (group S2). More samples have to be analyzed to corroborate this point.

4 Concluding remarks

Differences in chiral amino acid levels are clearly observed among the sample pools from different patients examined in this study, indicating that AD-related differences in CSF amino acid profiles can be examined effectively by chiral-MEKC-LIF. Some of the results shown in this work seem to rebut those shown in literature (e.g. not significant variation in D-Ser level was observed depending on the AD stage), while some of them agree with the results obtained by others (e.g. L-Asp occurs at significantly lower concentrations in AD CSF than normal CSF). Moreover, it is also preliminarily proposed that L-Glu, L-Arg, L-Asp, L-Lys and GABA could be used as biomarkers able to differentiate MCI patients who will progress to AD from those having a benign form of MCI without progression.

From this work, it is concluded that enantioselective procedures can open new perspectives in the discovery of biomarkers related to different stages of AD. However, more CSF samples will have to be analyzed in order to use L- and D-amino acid profiles from chiral-MEKC-LIF to make specific conclusions about neurological disorders such as AD. Moreover, the effect of other conditions (e.g. diet) on the level of D-amino acids found in CSF should also be studied, since the endogenous origin of some D-amino acids is already known (i.e. D-Ser), but the origin of others D-amino acids found in CSF samples in other works is not yet clear.

Figure 3. Scatter plot of the chiral-MEKC-LIF electropherograms (each point represents one electropherogram) in the plane defined by the two canonical variables (Root 1 and Root 2) obtained from FSDA selected variables: L-Glu, L-Arg, L-Asp, L-Lys and GABA. These amino acids classify successfully (100% of classification for all groups) the four pooled samples: S1 (control CSF), S2 (stable MCI), S3 (MCI with progression to AD) and S4 (AD). Ellipses surrounding the four groups are drawn for a better visualization. See text for more details.

5 References

[1] Wu, G. Y., *Amino Acids* 2009, *37*, 1–17.

[2] Arthur, R., *J. Comp. Med.* 2009, *8*, 46–61.

[3] Kirschner, D. L., Green, T. K., *J. Sep. Sci.* 2009, *32*, 2305–2318.

[4] Hamase, K., Morikawa, A., Etoh, S., *Anal. Sci.* 2009, *25*, 961–968.

[5] Hamase, K., Morikawa, A., Zaitsu, K., *J. Chromatogr. B* 2002, *781*, 73–91.

[6] Fonteh, A. N., Harrington, R. J., Tsai, A., Liao, P., Harrington, M. G., *Amino Acids* 2007, *32*, 213–224.

[7] Blennow, K., Hampel, H., Weiner, M., Zetterberg, H., *Nat. Rev. Neurol.* 2010, *6*, 131–144.

[8] Cedazo-Minguez, A., Winblad, B., *Exp. Gerontol.* 2010, *45*, 5–14.

[9] Gupta, V. B., Indi, S. S., Rao, K. S. J., *J. Mol. Neurosci.* 2008, *34*, 35–43.

[10] Tarbit, I., Perry, E. K., Perry, R. H., Blessed, G., Tomlinson, B. E., *J. Neurochem.* 1980, *35*, 1246–1249.

[11] Procter, A. W., Lowe, S. L., Palmer, A. M., Francis, P. T., Esiri, M. M., Stratmann, G. C., Najlerahim, A., Patel, A. J., Hunt, A., Bowen, D. M., *J. Neurol. Sci.* 1988, *84*, 125–140.

[12] Elison, D. W., Beal, M. F., Mazurek, M. F., Bird, E. D., Martin, J. B., *Ann. Neurol.* 1986, *20*, 616–621.

[13] Sasaki, H., Muromoto, O., Komzowa, I., Arai, H., Kosaka, K., Iizuka, R., *Ann. Neurol.* 1986, *19*, 263–269.

[14] Perry, E. K., Atack, J. R., Perry, R. H., Hardy, J. A., Dodd, P. R., Edwardson, J. A., Blessed, G., Tomlinson, B. E., Fairbairn, A. F., *J. Neurochem.* 1984, *42*, 1402–1410.

[15] Kaiser, E., Schoenknecht, P., Kassner, S., Hildebrandt, W., Kinscherf, R., Schroeder, J., *Neurodegener. Dis.* 2010, *7*, 251–259.

[16] Molina, J. A., Jiménez-Jiménez, F. J., Vargas, C., Gómez, P., de Bustos, F., Ortí-Pareja, M., Tallón-Barranco, A., Benito-León, J., Arenas, J., Enríquez-de-Salamanca, R., *J. Neural Transm.* 1998, *105*, 279–286.

[17] D'Aniello, A., Fisher, G., Migliaccio, N., Cammisa, G., D'Aniello, E., Spinelli, P., *Neurosci. Lett.* 2005, *388*, 49–53.

[18] Mochizuki, Y., Oishi, M., Hara, M., Takasu, T., *Annals Clin. Lab. Sci.* 1996, *26*, 275–278.

[19] Martinez, M., Frank, A., Dieztejedor, E., Hernanz, A., *J. Neural Transm.* 1993, *6*, 1–9.

[20] Bergquist, J., Gilman, D., Ewing, A. G., Ekman, R., *Anal. Chem.* 1994, *66*, 3512–3518.

[21] Thorsen, G., Bergquist, J., *J. Chromatgr. B* 2000, *745*, 389–397.

[22] Deng, Y. H., Wang, H., Zhang, H. S., *J. Sep. Sci.* 2008, *31*, 3088–3097.

[23] Lu, M. J., Chiu, T. C., Chang, P. L., Hoc, H. T., Chang, H. T., *Anal. Chim. Acta* 2005, *538*, 143–150.

[24] Herrero, M., Simó, C., García-Cañas, V., Fanali, S., Cifuentes, A., *Electrophoresis.* 2010, *31*, 2106–2114.

[25] Fisher, G. H., Petrucelli, L., Gardner, C., Emory, C., Frey, W. H., Amaducci, L., Sorbi, S., Sorrentino, G., Borghi, M., D'Aniello, A., *Mol. Chem. Neuropathol.* 1994, *23*, 115–124.

[26] Fisher, G., Lorenzo, N., Abe, H., Fujita, E., Frey, W. H., Emory, C., Fiore, M. M. D., D'Aniello, A., *Amino Acids* 1998, *15*, 263–269.

[27] Simó, C., Barbas, C., Cifuentes, A., *J. Agric. Food Chem.* 2002, *50*, 5288–5293.

[28] Mothet, J. P., *Pathol. Biol.* 2001, *49*, 655–659.

[29] Barañano, D. E., Ferris, C. D., Snyder, S. H., *Trends Neurosci.* 2001, *24*, 99–106.

[30] Blennow, K., *J. Int. Med.* 2004, *256*, 224–234.

[31] Rawls, S. M., Gomez, T., Stagliano, G. W., Raffa, R. B., *J. Pharmacol. Toxicol. Methods* 2006, *53*, 291–295.

[32] Boyd, B. W., Witowski, S. R., Kennedy, R. T., *Anal. Chem.* 2000, *72*, 865–871.

[33] Piepponen, T. P., Skujins, A., *J. Chromatogr. B* 2001, *757*, 277–283.

[34] Shaha, A. J., Crespib, F., Heidbrederb, C., *J. Chromatogr. B* 2002, *781*, 151–163.

[35] Lanctot, K. L., Herrmann, N., Mazzotta, P., Khan, L. R., Ingber, N., *Can. J. Psychiatry* 2004, *49*, 439–453.

[36] Chouinard, M. L., Gaitan, D., Wood, P. L., *J. Neurochem.* 1993, *61*, 1561–1564.

[37] Nagata, Y., Borghi, M., Fisher, G. H., D'Aniello, A., *Brain Res. Bull.* 1995, *38*, 181–183.

[38] Fuchs, S.A., de Sain-van der Velden, M. G. M., de Barse, M. M. J., Roeleveld, M. W., Hendriks, M., Dorland, L., Klomp, L. W. J., Berger, R., de Koning, T., *J. Clin. Chem.* 2008, *54*, 1443–1450.

[39] Albina, J. E., Mills, C. D., Henry, W. L., Caldwell, M. D., *J. Immunol.* 1990, *144*, 3877–3880.

[40] Jenkinson, C. P., Grody, W. W., Cederbaum, S. D., *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 1996, *114*, 107–132.

[41] Vural, H., Sirin, B., Yilmaz, N., Eren, I., Delibas, N., *Biol. Trace Elem. Res.* 2009, *129*, 58–64.

[42] Ishigami, A., Maruyama, N., *Geriatr. Gerontol. Int.* 2010, *10*, 53–58.

4.4. CAPÍTULO 6 (Chapter 6)

Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry Metabolomics

analytical chemistry

Toward a Predictive Model of Alzheimer's Disease Progression Using Capillary Electrophoresis−Mass Spectrometry Metabolomics

Clara Ibáñez,[†] Carolina Simó,[†] Pedro J. Martín-Álvarez,[†] Miia Kivipelto,[‡] Bengt Winblad,[‡] Angel Cedazo-Mínguez,[‡] and Alejandro Cifuentes*,[†]

[†]Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain

[‡]Karolinska Institute, NVS Department, KI-Alzheimer's Disease Research Center, 14186 Stockholm, Sweden

Ⓢ *Supporting Information*

ABSTRACT: Alzheimer's disease (AD) is the most prevalent form of dementia with an estimated worldwide prevalence of over 30 million people, and its incidence is expected to increase dramatically with an increasing elderly population. Up until now, cerebrospinal fluid (CSF) has been the preferred sample to investigate central nervous system (CNS) disorders since its composition is directly related to metabolite production in the brain. In this work, a nontargeted metabolomic approach based on capillary electrophoresis−mass spectrometry (CE−MS) is developed to examine metabolic differences in CSF samples from subjects with different cognitive status related to AD progression. To do this, CSF samples from 85 subjects were obtained from patients with (i) subjective cognitive impairment (SCI, i.e. control group), (ii) mild cognitive impairment (MCI) which remained stable after a follow-up period of 2 years, (iii) MCI which progressed to AD within a 2-year time after the initial MCI diagnostic and, (iv) diagnosed AD. A prediction model for AD progression using multivariate statistical analysis based on CE−MS metabolomics of CSF samples was obtained using 73 CSF samples. Using our model, we were able to correctly classify 97−100% of the samples in the diagnostic groups. The prediction power was confirmed in a blind small test set of 12 CSF samples, reaching a 83% of diagnostic accuracy. The obtained predictive values were higher than those reported with classical CSF AD biomarkers (Aβ42 and tau) but need to be confirmed in larger samples cohorts. Choline, dimethylarginine, arginine, valine, proline, serine, histidine, creatine, carnitine, and suberylglycine were identified as possible disease progression biomarkers. Our results suggest that CE−MS metabolomics of CSF samples can be a useful tool to predict AD progression.
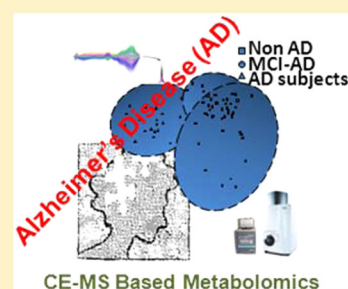
Alzheimer's disease (AD) is characterized by progressive loss of memory and other cognitive functions leading to dementia. The long duration of AD and its increased incidence with age constitutes a large emotional and financial weight for patients, their families, and society. It is predicted that the worldwide number of AD cases, presently about 36 million, will triple by 2050.[1] Thus, it is of extreme importance to solve the most important AD questions raised, i.e., origin, causes, treatment, prevention, and early and accurate diagnosis. Most AD biomarkers deeply studied so far are protein molecules. Thus, increased total tau (t-tau) and phospho-tau (p-tau) while decreased amyloid beta (Aβ) levels have been observed in AD subjects, in comparison with nondemented subjects.[2] Sensitivity and specificity values of 80−88% are obtained in the diagnosis of advanced cases of AD when these biomarkers are combined.[3,4] On the other hand, advanced medical brain imaging techniques (computed tomography, nuclear magnetic resonance imaging, and single photon or positron emission computed tomography) may help to diagnose the existence of dementia but not specific dementia due to AD.[5] The combination of the analysis of specific protein levels and brain imaging might increase AD diagnosis up to 90%,[4] but the economic cost per patient arises immensely (∼6 000−10 000 euros approximately per patient).

In AD research, prediction of progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD) is also of major interest. Moreover, there is a clear interest to search for AD biomarkers in biofluids. Up until now, cerebrospinal fluid (CSF) has been the preferred sample for the discovery of new biochemical insights into CNS disorders since its composition is directly related to metabolite production in the brain.[6,7] The role of CSF includes a mechanical protection of the central nervous system (CNS) against trauma and transport of nutrients to ensure the homeostasis of CNS cells.[8] Nowadays, metabolomics is playing an increasingly important role in the correlation of the biofluid/tissue metabolome and disease. At present the human CSF metabolome information available in HMDB contains 468 identified metabolites.[9] In AD research, nontargeted metabolomics of human CSF has been tackled by few researchers. The changes in the CSF metabolic profiles by gas chromatography/mass spectrometry (GC/MS) and liquid chromatography−mass spectrometry (LC−MS) from 79 AD patients and 51 healthy controls has been presented in a recent

Analytical Chemistry　　　　　　　　　　　　　　　　　　　　Article

work.[10] Cysteine, uridine, and cortisol, were determined as significant metabolites allowing distinction between AD and healthy subjects.[10] A cation exchange nano-LC/MS method has been also used to detect 55 unidentified hydrophilic compounds that differentiate AD from healthy patients.[11] Free amino acid and dipeptide profiles in CSF were observed to be different in plasma and urine of healthy and AD patients.[12] More recently, a chiral micellar electrokinetic chromatography with laser-induced fluorescence detection was developed to investigate D- and L-amino acid contents in cerebrospinal fluid (CSF) samples related to AD.[13]

The physicochemical complexity and the large dynamic range of metabolites in biofluids/tissues require the application of multiple analytical platforms for a global metabolome coverage. In this regard, besides the possible AD biomarkers mentioned above, the works published so far on the use of LC−MS and GC/MS for AD metabolomics have mainly focused on the lipid fraction,[14] bile acids family,[15] or sphingolipidome.[16] Therefore, the use of capillary electrophoresis−mass spectrometry (CE−MS) can provide very useful information in a fast way about ionic and highly polar metabolites that cannot be easily obtained by either LC−MS or GC−MS methods.[17] Thus, nontargeted CE−MS metabolomic approaches could potentially provide a variety of new biomarkers to monitor AD. In this work, a nontargeted metabolomic approach based on CE−MS is developed to examine metabolic differences in CSF samples from subjects with different cognitive status related to AD progression. After CE−MS metabolomic analysis, multivariate statistical analysis was applied for the construction of a prediction model corroborating the huge potential of this methodology. To our knowledge, this is the first time that CE−MS metabolomics is applied to predict the progression of AD in nondiagnosed AD patients based on CSF profiles.

■ EXPERIMENTAL SECTION

Chemicals. All chemicals were of analytical reagent grade, except reagents and solvents employed in the preparation of CE electrolytes and sheath liquid that were of MS grade, and all used as received. Detailed description is given as Supporting Information.

Subjects. Subjects were patients who attended the Karolinska University Hospital and received an initial diagnosis of subjective cognitive impairment (SCI), mild cognitive impairment (MCI), or Alzheimer's disease (AD). After a follow-up period of 2 years, a subset of MCI individuals developed AD and the rest of the patients remained stable after the initial diagnosis. The present study recruited a total number of 85 patients (see the Supporting Information). After CSF sampling, the samples were randomly split into a training set for predictor discovery and classification training, and a test set for class prediction of blinded samples to evaluate the chosen predictors and parameters (Table S-1, Supporting Information). Thus, the training set of 73 samples presented the following clinical diagnostics: 19 subjects with SCI (SCI-nonAD group), 22 subjects with MCI which remained stable after a follow-up period of 2 years (MCI-nonAd group), 9 subjects with MCI which progressed to AD within a 2-year time after the initial MCI diagnostic (MCI-AD group) and 23 subjects with AD (AD group). The test set of 12 blinded samples presented the following clinical diagnostic: 4 with SCI, 2 with MCI which remained stable after a follow-up period of 2 years, 4 with MCI which progressed to AD within 2-year time

after the initial MCI diagnostic, and 2 with AD. Demographic characteristics and clinical diagnosis of studied subjects are summarized in Table S-1 in the Supporting Information and detailed in Table S-2 in the Supporting Information.

CSF Sampling and Metabolite Extraction. CSF was collected by lumbar puncture as described in the Supporting Information. The metabolite extraction method was employed as described previously[17] with minor modifications. Briefly, internal standards were added to the CSF (100 μL). CSFs were then ultrafiltrated by using 3 kDa Amicon Ultra 0.5 mL centrifugal devices from Millipore (Billerica, MA). The centrifugation was performed at 14 000g for 40 min at 20 °C. Filtered fractions (with less than 3 kDa compounds) were collected and stored at −80 °C until CE−time-of flight (TOF) MS analysis.

CE−TOF MS Analysis. CE analyses were carried out in a P/ACE 5500 CE apparatus from Beckman Instruments (Fullerton, CA). The instrument was controlled by a PC running the System Gold software from Beckman. Poly(vinyl alcohol) (PVA) coated capillary with 50 μm i.d. and 125 cm effective length (from Agilent Technologies, Palo Alto, CA) and uncoated fused-silica capillaries (50 μm i.d. and 87 cm total length) from Composite Metal Services (Worcester, England) were coupled to the mass spectrometer through an orthogonal electrospray ionization (ESI) interface model G1607A from Agilent Technologies (Palo Alto, CA). Electrical contact at the electrospray needle tip was established via a sheath liquid. A TOF MS instrument (micrOTOF) from Bruker Daltonics (Bremen, Germany) was employed. The instrument was controlled by a PC running the micrOTOF control software from Bruker Daltonics. Injections were made at the anodic end using an N_2 pressure of 0.5 psi (34.5 mbar) for 80 s. The electrophoretic separation was achieved applying +25 kV as running voltage at a constant temperature of 25 °C. The micrOTOF was operated to acquire spectra in the m/z range of 50−500. Each sample was analyzed in duplicate by CE−MS, and the capillary was conditioned with the separation buffer during 4 min between runs. Metabolic profiles were obtained in less than 15 min. Other conditions are given as Supporting Information.

Data Processing and Multivariate Statistical Analysis. Raw data signals obtained from CE−MS experiments were exported to the open source MS exchange format netCDF using Data Analysis 4.0 (Bruker). Automatic peak detection, migration time alignment, and peak integration was performed using MZmine software (version 2.2). The resulting metabolic feature list was then exported as a csv file and opened using Microsoft Office Excel 2010 for calculation of mean values from technical duplicates and for peak area normalization considering the sum of the corrected peak areas of the two internal standards. A deeper description of the data processing procedure is given in the Supporting Information and Table S-3. The final data table with the normalized areas of the 160 metabolites analyzed in the 85 samples (73 in the training set and 12 in the test set) was used for subsequent statistical treatment and was imported into the statistical software.

The statistical methods used for data analysis were principal component analysis (PCA) from standardized variables, to detect possible outliers in each of the four groups of samples belonging to training set (SCI-nonAD, MCI-nonAd, MCI-AD, and AD groups); linear discriminant analysis (LDA), with forward stepwise procedure, to select the variables most useful in differentiating the four groups of samples; canonical variate

analysis (CVA) to obtain a low-dimensional graphical representation of the samples that separates the groups as much as possible; and leave-one-out cross-validation (LOO−CV) procedure to assess the predictive power of the LDA classification method.[18,19] STATISTICA (version 7, Statsoft, *Tulsa, OK, www.statsoft.com*) and SPSS (*version 19, IBM, Chicago, IL, www.spss.com*) programs for Windows were used for data processing and were run on a PC Intel Core i7-2600K processor (8 M cache, up to 3.80 GHz) with 64 bits operating under Microsoft Windows 7.

Significant metabolic features selected by forward stepwise LDA classification model were identified by matching the experimental and theoretical *m/z* values contained in the Human Metabolome Database[9] and METLIN[20] within a mass accuracy window of 15 ppm. The isotopic pattern was also considered to support the putative identification. Standards were then used to confirm metabolite identification by means of sample coinjection and subsequent CE−MS analysis.

■ RESULTS AND DISCUSSION

Analysis of CSF Samples: CE−MS Method Development. CE is particularly suited for the rapid separation of ionic, weakly ionic, and/or highly polar metabolites with very high resolution using extremely small volumes of reagents and sample. In this work, different analytical conditions were tested for an adequate CE−MS analysis of the CSF samples. Direct CE−MS analysis of CSF samples was not possible due to electrical current instability and large migration times variation probably induced by protein adsorption onto the capillary wall, and these problems were solved by including the filtration step as indicated in CSF Sampling and Metabolite Extraction. CE−MS conditions were optimized in terms of signal stability, number of detected features (metabolites), resolution, and sensitivity. On the basis of our previous experience with CE−MS for other metabolomic studies,[17] different injection volumes of the sample (40, 50, 60, 80, and 100 s) at 0.5 psi (34.5 mbar) were evaluated. Injection times above 100 s negatively affected the resolution, while with injection times shorter than 80 s, lower peak intensity (and therefore a lower number of features) was observed. Different background electrolytes (BGE) compositions were tested, from 0.25 to 1.5 M acetic and formic acid. It was observed that formic acid provided more peaks and higher signal/noise ratios than acetic acid. Besides, 0.5 M formic acid concentration showed the best electrophoretic conditions in terms of resolution. Moreover, two different types of capillary were compared, namely, a PVA coated capillary and a bare-fused silica capillary. Stability of the CE current was higher when a bare-fused silica capillary was used, probably due to the presence of a residual electroosmotic flow (EOF) that helped to maintain the electrical contact at the end of the capillary. Overall, a higher number of features was obtained when the uncoated capillary was used. Sheath liquid compositions of 25/75, 50/50, 75/25, and 100/0 for both methanol/water and isopropanol/water were tested. Signal/noise ratio and electrospray stability were better with 50/50 isopropanol/water. Then, sheath liquid flow was optimized from 2 to 5 µL/min, and optimum electrospray conditions were obtained at 4 µL/min. Nebulizer gas pressure was varied from 0.3 to 0.5 bar obtaining a better intensity signal without loss of resolution at 0.4 bar.

The best results were obtained under the following conditions: bare fused silica capillary containing 0.5 M formic acid at pH 1.8 as BGE; injection at the anodic end using N_2

pressure of 0.5 psi (34.5 mbar) for 80 s; running voltage of 25 kV; isopropanol/water (50/50, v/v) as the sheath liquid delivered at a flow rate of 4 µL/min and using nebulizer and drying gas flows of 0.4 bar N_2 and 4 L/min N_2, respectively. Under these acidic separation conditions, most compounds are positively charged and migrate as cations. An example of the CE−MS electropherograms obtained is given in Figure 1 in
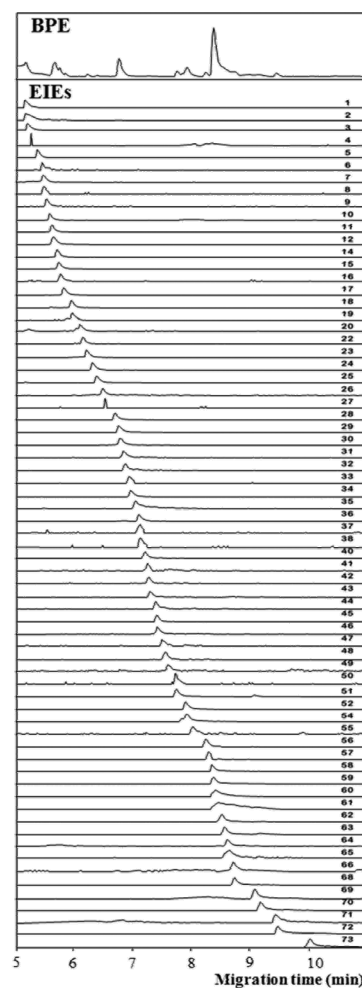
Figure 1. CE−electrospray (ESI)-TOF MS base peak electropherogram (BPE) of metabolites in CSF samples and extracted ion electropherograms (EIE) of the main metabolites found in CSF of a randomly selected sample. For metabolite identifications see Table S-5 in the Supporting Information. Other CE−MS conditions are described in the Experimental Section.

which the base peak electropherogram and some selected extracted ion electropherograms obtained from a CSF sample are shown. CSF samples belonging to different patients presented similar CE−MS profiles. Also, some electrophoretic comigrations of close metabolites were observed under the selected CE separation conditions. The use of high-resolution mass analyzers, like the one used in this work (time-of-flight, TOF-MS), is critical to obtain not only accurate mass determination but also to carry out the CE−MS analysis of

closely eluting compounds. The 85 CSF samples described in Table S-1 in the Supporting Information were analyzed in duplicate under the selected CE−MS conditions. Before data processing, visual inspection of all obtained electropherograms was done in order to confirm the correct CE−MS analysis of all injected samples.

Data Treatment and Prediction Model Based on CE−MS Metabolomics. MZmine analysis of the data obtained from the CE−MS analysis of the 85 samples (injected in duplicate) was then carried out. MZmine provided a final matrix for each sample containing 160 time-aligned metabolic features with their corresponding migration time, m/z, and peak area. The matrix with the mean values of peak area from the duplicates was normalized. These peak areas from CE−MS metabolomics of 73 CSF samples as the training set and 12 CSF samples as the test set (see Table S-1 in the Supporting Information) were subjected to the following statistical analysis: PCA was first applied to each group of samples from the training data set (i.e., 73 CSF samples, see Table S-1 in the Supporting Information) in order to detect possible outliers before applying the forward stepwise LDA classification method. In Figure S-1 in the Supporting Information, samples from SCI-nonAD (Figure S-1A in the Supporting Information), MCI-nonAd (Figure S-1B in the Supporting Information), MCI-AD (Figure S-1C in the Supporting Information), and AD (Figure S-1D in the Supporting Information) groups are plotted in the plane defined by the first two principal components (PC). The total variance explained by the first two PC in each case was 29.6% for SCI-nonAd group, 36.1% for MCI-nonAd group, 51.9% for MCI-AD group, and 41% for AD group. Two samples were considered outliers (namely, samples 108 and 17 from MCI-nonAd and AD patients, respectively) as they were placed out of the 99% confidence ellipse in the PCA graphical representation (see Figure S-1 in the Supporting Information) and were excluded for the subsequent statistical analysis. Patient 108 did not have any other clinical feature mentionable, while patient 17 was also diagnosed with hypothyroidism, hypertension, and ischemic heart disease. Forward stepwise LDA procedures (with probabilities values set at 0.05 to enter and to remove variables) were applied to the training data set in order to select the metabolites that most differentiate the four groups of samples (SCI-nonAd, MCI-nonAd, MCI-AD, and AD). The results obtained of the 73 CSF samples from the training data set, by forward stepwise LDA as well as the quality indicators (sensitivity, specificity, and positive and negative predictive satisfactory values) are shown in Table S-4 in the Supporting Information. A total of 10 metabolites (IDs 7, 3, 25, 65, 48, 43, 15, 1, 70 and 2, see Table S-5 in the Supporting Information) were selected. A 90.1% correct classification for the four-group comparison was achieved with these selected metabolites by LDA. More precisely, 94.7% of samples in SCI-nonAd group were correctly classified (1 sample out of 19 was misclassified as AD), 85.7% of samples in MCI-nonAd group were correctly classified (3 samples out of 22 were misclassified as SCI-nonAd), 88.9% of samples in MCI-AD group were correctly classified (1 sample out of 9 was misclassified as MCI-nonAd), and 90.9% of samples in AD group were correctly classified (2 samples out of 22 were misclassified in SCI-nonAd and MCI-nonAd groups).

In order to assess the predictive power of the LDA method with these 10 selected metabolites, LOO−CV was used. Thus, LOO−CV classification was made by holding each observation

out (one-at-a-time) and building a LDA model from the remaining training observations and then classifying the held out observation using this model. The percentages of correct classification of the samples using the LOO−CV procedure are given in Table S-4 in the Supporting Information. As can be seen and as expected, a lower correct assignment percentage of 74.6% for the total number of samples (among the four groups of samples) was obtained. Accordingly, 78.9% of samples in SCI-nonAd were correctly classified (4 out of 19 were misclassified: 3 as MCI-nonAd and 1 as AD), 71.4% of samples in MCI-nonAd were correctly classified (4 out of 21 were misclassified: 2 as SCI-nonAd, 1 as MCI-AD and 1 as AD), 55.6% of samples in MCI-AD were correctly classified (4 out of 9 were misclassified: 1 as MCI-nonAd and 3 as AD), and 81.8% of samples in AD group were correctly classified (4 out of 22 were misclassified: 2 as SCI-nonAd, 1 as MCI-nonAd, and 1 as MCI-AD).

Forward stepwise LDA (with probability values of 0.05 to enter and remove variables) and LOO−CV procedures were also applied to select the variables most useful to differentiate two groups of samples (see Table S-4 in the Supporting Information). In each particular case, different metabolic features were selected and higher correct classification percentages were achieved either by the standard LDA or LOO−CV procedures applied to the selected metabolites. Calculated sensitivity, specificity, positive and negative predictive percentages values obtained from LDA and LOO−CV procedures are also included in Table S-4 in the Supporting Information. Thus, a 95% of correct classification between SCI-nonAd and MCI-nonAd groups was obtained considering eight metabolites (IDs 3, 2, 52, 4, 10, 58, 56, 17) by the standard LDA method, and a 80% correct assignment with the LOO−CV procedure. Values of 95.2%, 94.7%, 95.2%, and 94.7% for sensitivity, specificity, positive and negative predictive values, respectively, were obtained from the classification results with the LDA method. Percentages of 71.4%, 89.5%, 88.2%, and 73.9% for sensitivity, specificity, positive and negative predictive values, respectively, were obtained by the LOO−CV procedure. A 97.5% of correct classification between groups SCI-nonAd and AD (1 out of 22 AD samples was misclassified as SCI-nonAd) with five selected metabolites (IDs 25, 65, 12, 2, 30) by LDA and 97.6% by the LOO−CV procedure (only 1 out of 22 AD samples was misclassified as SCI-nonAd). Satisfactory percentages (above 80%) for sensitivity, specificity, positive and negative predictive values (100%, 94.7%, 95.5%, and 100%, respectively) by the standard LDA method or LOO−CV procedure (95.5, 100, 100, and 95%, respectively) were obtained. For the rest of two-group comparisons, 100% of correct classification by LDA was achieved. LOO−CV was performed in parallel to LDA techniques, and slightly lower values of correct classification were obtained when comparing only one group with another one (detailed information is also given in Table S-4 in the Supporting Information). In all these two-group comparisons, satisfactory percentages of correct classification by LOO−CV (above 80%) for sensitivity, specificity, positive and negative predictive values were obtained. The lowest percentage of correct classification (95% for LDA procedure and 80% by LOO−CV) and values below 80% for sensitivity and negative predictive value were observed when SCI-nonAd and MCI-nonAd groups were compared, suggesting a higher grade of similarity in terms of metabolic profile. Actually, these two groups of subjects did not develop AD within 2-year time after the first clinical diagnosis.

Following this finding and with the purpose of improving the classification power of the LDA models based on CE−TOF MS metabolome analysis, a further statistical analysis merging SCI-nonAd and MCI-nonAd in a single group (hereinafter called, nonAd) was carried out. Thus, PCA was first applied to this nonAd group to detect possible outliers. In Figure S-2 in the Supporting Information, the samples in the non-AD group are plotted in the plane defined by the first two PC, which explained 28.9% of the total variance. Sample 108 placed out of the 99% confidence ellipse was considered an outlier and it was excluded for the subsequent statistical analysis. After the exclusion of the same two outliers (samples 108 and 17 from non-AD and AD patients, respectively) the forward stepwise LDA method (with probabilities values of 0.05 to enter and remove variables) was applied to the 71 samples, to select the metabolites most useful to differentiate the three categories of CSF samples: non-AD, MCI-AD, and AD. Detailed group assignments by LDA and LOO−CV as well as the quality indicators (sensitivity, specificity, and positive and negative predictive satisfactory values) are given in Table 1. As can be seen in Table 1, a better classification of 97.2% was obtained in this case for the three-groups comparison, selecting 14 metabolites (IDs 7, 48, 15, 43, 32, 46, 35, 25, 3, 71, 21, 29, 52, 17). Besides, 100% of samples from non-AD group were correctly classified, while only 2 samples (1 out of 9 in group MCI-AD and 1 out of 22 in group AD) were misclassified as non-AD group. In order to assess the predictive power of the LDA classification method with these 14 selected metabolites, the LOO−CV procedure was used and the percentages of correct classification of the samples are given in Table 1. Thus, an 87.3% correct assignment of the total of samples was obtained (95% for non-AD group, 55.6% for MCI-AD, and 86.4% for AD group).

Considering the 14 selected metabolites from the forward stepwise LDA procedure, a CVA was applied to the 71 samples of the training data set. In Figure 2, the 71 samples from the training data set are plotted in the plane defined by the two calculated canonical variables. Also, 95% confidence canonical ellipses are included showing well separated groups with slight overlapping. The non-AD and AD patients are well separated mainly by the first canonical variable (CV1). Further investigation based in the obtained factor structure matrix (with correlations between the selected metabolites and the canonical variables) showed that the discrimination by CV1 was due to significant changes in metabolites 48, 32, and 25. A secondary difference between AD patients and MCI-AD was also detected by the second canonical variable (CV2). In this case, and from the factor structure matrix, changes in metabolites 52 and 7 were especially important to achieve this discrimination. This difference is particularly interesting due to the huge interest that nowadays exists in discriminating MCI-AD from AD subjects. Thus, an early pharmacological/psychological treatment could be ideally applied to these identified patients to prevent or delay the AD symptoms, maintaining in this way a better quality of life for a longer time.

Usefulness of CE−MS Metabolomics: Prediction Model Application. To corroborate the usefulness of the CE−MS metabolomics methodology, the latter LDA model was applied to classify the 12 blinded CSF samples from the test set (see Table S-1 in the Supporting Information). Detailed classification results of these samples are given in Table 2, and a CVA graphical representation of this classification is shown in Figure 2. By using this methodology, 83% of the test samples

Table 1. Three-Group Comparison Assignments by LDA and LOO−CV

group comparisons[a]	no. of selected metabolites for the classification[b]	LDA					LOO−CV				
		% correct classification	% sensitivity	% specificity	% positive predictive value	% negative predictive value	% correct classification	% sensitivity	% specificity	% positive predictive value	% negative predictive value
non-AD vs MCI-AD vs AD	14 (7[c], 48, 15, 43, 32, 46, 35, 25, 3, 71, 21, 29, 52, 17)	97.2% 100% (non-AD), 88.89% (MCI-AD), 95.4% (AD)					87.3% 95% (non-AD), 55.6% (MCI-AD), 86.4% (AD)				
non-AD vs MCI-AD	8 (7, 65, 44, 25, 2, 1, 3, 29)	100% 100% (non-AD), 100% (MCI-AD)	100	100	100	100	95.9% 97.5% (non-AD), 88.9% (MCI-AD)	88.9	97.5	88.9	97.5
MCI-AD vs AD	19 (60, 48[c], 32, 7[c], 11, 71[c], 29[c], 12[c], 10[c], 15, 50, 2, 41, 67, 53[c], 59, 17, 1[c], 38)	100% 100% (MCI-AD), 100% (AD)	100	100	100	100	100% 100% (MCI-AD), 100% (AD)	100	100	100	100
non-AD vs AD	11 (48, 65, 25, 7[c], 15, 36, 31, 32, 35, 53, 43)	98.4% 100% (non-AD), 95.4% (AD)	95.5	100	100	97.6	93.5% 95% (non-AD), 90.9% (AD)	90.9	95	90.9	95

[a]Sample groups compared in each case are marked in the first column. Calculated sensitivity, specificity, and positive and negative predictive values are also given for both classification procedures. [b]In parentheses is the ID of significant metabolites. [c]More than one compound (adduct, complex...) corresponding to the same m/z value (the same metabolite). Metabolite IDs order are given according to the forward stepwise LDA significance.

163
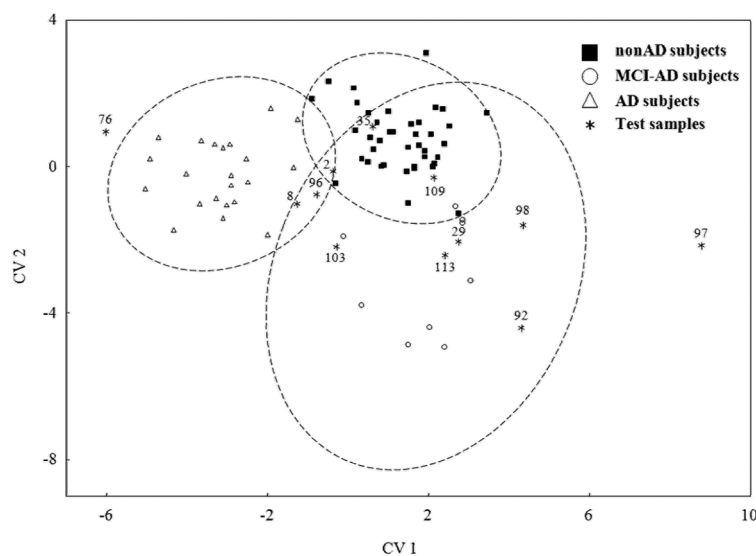
Figure 2. Plot for the training and test samples in the plane defined by the two canonical variables (CV1 and CV2) obtained by CVA considering the 14 selected highlighted metabolites after forward stepwise LDA. The 95% canonical ellipses are also included. Non-AD, MCI-AD, and AD CSF samples are represented by black squares, white circles, and white triangles, respectively. The 12 samples from the test set are represented by asterisks and their ID number.

Table 2. Classification of the CSF Samples from the Test Set According to the Conventional Medical Diagnostic vs the CE−MS-Based Metabolomic Method of This Work

patient ID	medical diagnostic[a]	CE−MS method
02	non-AD	non-AD
35	non-AD	non-AD
96	non-AD	non-AD
109	non-AD	non-AD
29	non-AD	MCI-AD
92	non-AD	MCI-AD
97[b]	MCI-AD	MCI-AD
98	MCI-AD	MCI-AD
103	MCI-AD	MCI-AD
113	MCI-AD	MCI-AD
08	AD	AD
76	AD	AD

[a]Classification performed at the Karolinska Institute (see the Experimental Section). [b]High Mahalanobis distance.

were correctly assigned to their corresponding group, being two samples apparently misclassified. Samples incorrectly assigned corresponded to samples 29 and 92, both from patients classified as MCI-AD according to our CE−MS method, while according to the medical diagnostic they belong to the same MCI-nonAd group. Patient 92 with MCI at the time of CSF sampling did not develop AD after a follow-up period of 2 years. This in an unusual case since this patient is a young man (only 51 years old) in contrast to other patients participating in the study (more than 60 years old on average). Thus, a longer follow-up period would be expected to be needed to diagnose (or not) this patient with AD. On the other hand, patient 29 presented genotype ApoE $\varepsilon3/\varepsilon4$. The presence of the e4 allele increases the risk to develop AD, accelerating the deposition of Aβ 42 and, consequently, decreasing free Aβ42 in CSF.[21,22] Although this subject did not develop AD in the followed time,

abnormal low values of Aβ 42 typically found in AD patients were found in this patient (545 pg/mL). Moreover, high t-tau and p-tau levels, also typical from AD patients, were abnormally detected (t-tau 550 pg/mL and p-tau 84 pg/mL) in this subject. Another interesting case was sample 97 from a subject with MCI which progressed to AD within the 2-year time after the first MCI diagnosis. Our proposed CE−MS method correctly classified this sample; however, a large Mahalanobis distance to the centroid inside its group was observed, and its location was out of the 95% confidence ellipse in the CVA graphical representation (Figure 2). This patient suffered from concomitant diseases, namely, AD and polio disease, which could be the reason for differences in the metabolic profile. Indeed, it has been stated that the polio virus can infect and replicate in motor neurons destroying them,[23] which could explain the results obtained.

The established markers for AD and CSF biomarkers are total tau (tTau), hyperphosphorylated tau (p-Tau), and the 42 amino acid form of Aβ-42. In a very recent multicenter cohort study[24] using more than 1500 CSF samples, with similar subclassification than ours (SCI, MCI-stable, MCI-AD, and AD), a positive predictive value of 57% in 64−74 year-old patients and a 67% in individuals older than 75 years was reported. The negative predictive values were 79% and 90%, respectively. They used a formula that combines Aβ-42 and p-Tau measurements. Our model based on CE−MS metab-olomics could obtain better prognostic values. However, to confirm this result, a larger study including samples from different centers seems necessary. Since the pathological processes in neurodegenerative disorders are multiple, the diagnostic accuracy may be further improved by combining several markers. The sensitivity and specificity of Aβ-42 and p-Tau could be sufficient to be used for screening but not as an early diagnostic test. The inclusion of other molecules and the stratification of the sample could improve the power of diagnostics. For example, we reported that the t-tau/

Analytical Chemistry
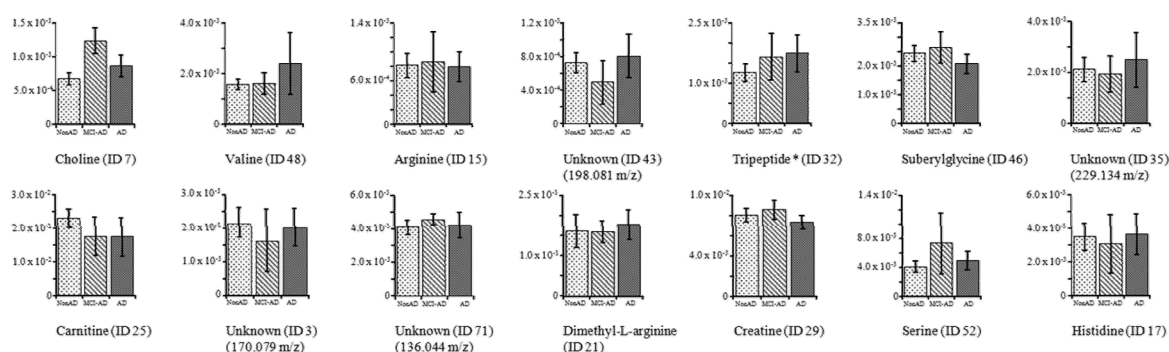
Figure 3. Bar plots and 95% confidence interval of the mean CE−MS peak areas of the 14 metabolites that come out as key compounds after the forward stepwise LDA classification procedure.

(Aβ42*insulin) ratio showed a striking better sensitivity and specificity than the total-tau/Aβ1-42 ratio for early AD diagnosis in women.[25] The combination of metabolite analysis with disease-related biomarkers has not been fully explored, and it could serve for improving early diagnosis.

CE−MS Metabolomics Results: Biological Meaning and Potential Biomarkers. For a better understanding of metabolic changes in the early stages of AD, tentative identification of significant metabolites was carried out (Table S-5 in the Supporting Information shows the identification of the metabolites that came out as significant ones during the development of the LDA classification methods discussed above). As already stated for the three-group LDA prediction model (non-AD vs MCI-AD vs AD), a total number of 14 significant metabolites were obtained (namely, IDs 7, 48, 15, 43, 32, 46, 35, 25, 3, 71, 21, 29, 52, and 17). Identification of these metabolites was carried out following the approach described in Data Processing and Multivariate Statistical Analysis, and the results obtained are shown in Table S-5 in the Supporting Information. When available, standards were used to confirm metabolite identity. Figure S-3 in the Supporting Information shows the CE−MS results (given as extracted ion electropherograms, EIE) obtained from the coinjection of standards with a CSF sample. Following this procedure, the existence of choline, arginine, histidine, dimethylarginine, carnitine, creatine, valine, serine, and proline was confirmed in CSF samples. This procedure also allowed us to discard some metabolites that could have been wrongly assigned based only on their molecular mass. This is the case of metabolites 3, 35, 43, and 71 with $[M + H]^+$ values of 170.079, 229.134, 198.081, and 136.044, respectively. Initially, they were tentatively identified as norepinephrine (M_w = 169.073 898), leucine-proline (M_w = 228.147 385), L-dopa (M_w = 197.068 802), and homocysteine (M_w = 135.035 400). However, the migration time of the coinjected standards was completely different from expected, and therefore, these metabolites still remain unknown.

In Figure 3, bar plots corresponding to CE−MS mean values and the 95% confidence interval of peak areas from the 14 selected metabolites by forward stepwise LDA are presented. In total, 5 out of the 14 metabolites showed significant differences ($p < 0.05$) in the mean values of their peak areas when the least significant differences (LSD) test was performed. These metabolites were choline, with highest values in the MCI-AD subjects, lowest values in non-AD subjects, and intermediate values in AD subjects; valine, with higher values in the AD

subjects than in MCI-AD and non-AD subjects; a tripeptide with higher values in the AD subjects than in non-AD subjects; carnitine with higher values in the non-AD subjects than in AD subjects; serine with higher values in the MCI-AD subjects than in non-AD subjects. It is interesting to remark the relation already described in the literature for many of these possible biomarkers with AD. Choline is involved in betaine metabolism, methionine metabolism, and phospholipid biosynthesis, acting as a precursor of the neurotransmitter acetylcholine, and its presence at normal concentrations reflects adequate neurotransmitter activity and metabolism in the brain and CNS.[26] Increased amounts of choline and related metabolites (glycerophosphocholine and phosphocholine) have already been found in AD patients when compared with controls.[27] For some metabolites, this relation is contradictory as can be deduced from the works published on dimethylarginine, an endogenous inhibitor of nitric oxide synthase that is found in elevated levels in patients with vascular disease, especially suffering end-stage renal disease. Thus, in some works there is described some relation between dimethylarginine and AD,[28] while in another work the authors specifically mention that there is no relation between the levels of this metabolite in CSF and AD.[29] As can be seen in Figure 3, free amino acids valine, serine, histidine, and arginine showed diverse patterns; two of them increased in CSF of AD patients (e.g., valine and serine), histidine decreased in MCI-AD patients, while a slight decreasing of arginine was detected in CSF of AD patients in good agreement with a recent work published by Fonteh et al.[12] Arginine is involved in mechanisms of nitrogen detoxification, and alteration of these processes can be a consequence of an increase of amino acid metabolism demanding more protecting capacity and more detoxification (more arginine degradation) than would be expected with the loss of brain tissue.[13] As can be seen in Figure 3, creatine was lower in CSF of AD patients than in CSF from non-AD or MCI-AD patients, in good agreement with recent findings on the level of this compound determined at the hippocampus by magnetic resonance spectroscopy in AD patients.[30] Creatine plays a crucial role in energy provision, besides it has been recently demonstrated in animal models that creatine administration can act against neurodegeneration and aging.[31] Thus, because of its neuroprotective activity, lower concentrations of creatine are expected in AD patients,[32] while it has been suggested that creatine supplementation could attenuate the degenerative state induced by central nervous diseases including AD.[33] As can be seen in Figure 3, carnitine was lower in CSF from MIC-AD and

Analytical Chemistry

AD patients than in CSF from non-AD subjects. Carnitine, besides having a critical role in energy provision and balance via fatty acid and carbohydrate utilization, has been described as a neuroprotective compound that counteracts the brain damage induced by oxidative stress in AD as well as an essential compound in the modulation and promotion of synaptic neurotransmission, most notably cholinergic neurotransmission.[34] This work shows for the first time a possible relation between a metabolite as suberylglycine and AD. Suberylglycine is an acyl glycine. These compounds, i.e., acyl glycines, are minor metabolites of fatty acids. The measurement of these metabolites in body fluids can be used to diagnose disorders associated with mitochondrial fatty acid β-oxidation.

CONCLUSIONS

The diagnosis of AD in its early stage is still challenging and demands the development of new analytical strategies. The CE−MS metabolomics approach developed in this paper makes it possible to identify differences in the CSF metabolome from patients with different cognitive status related to AD progression. CE−MS analysis of CSF metabolome allowed the development of multivariate models for the efficient discrimination and classification of CSF samples related to different stages of AD progression. Despite the intrinsically biased results obtained by a single analytical platform, the presented metabolomic approach shows promising potential for early detection of AD. This is one of the main achievements of this work, i.e., the promising possibility of AD prediction prior to dementia onset based on CE−MS analysis of the CSF metabolome. The proposed biomarkers, obtained after multivariate statistical analysis of CE−MS metabolomic results, may improve individual prognostic or predictive interpretation of subjects presenting MCI. The consistency of the proposed biomarkers has to be demonstrated in a larger number of subjects (especially in the MCI-AD group), obtained from different centers, in order to improve the statistical classification confidence. Moreover, according to the typical protocols for AD treatment, once it is suspected that a patient has AD, it should be treated with acetylcholinesterase inhibitors, while additionally due to their age-related health problems, these patients are usually taking other drugs. Although for statistics, we have only considered metabolites that were present in all patients and not those present only in some of them; some drugs-induced noise to our results cannot be ruled out. Combination of marker molecules, classical and novel, should be also explored. Indeed, AD is not a homogeneous disorder, and uncovering different etiologies could be of great importance for a proper diagnosis, evaluation of disease progression, and adjusting future therapeutic strategies. Although metabolomics still remains challenging due to the structural diversity and broad dynamic concentration range of endogenous metabolites in biofluids, the widespread use of MS-based technologies can make this approach to be part of routine diagnosis of patients with cognitive impairment in the nondistant future. On the other hand, although lumbar puncture is widely spread, it does not satisfy the criteria of minimal invasiveness. Thus, future studies should focus on developing predictive tools of AD based on the metabolome of more accessible samples (e.g., blood, urine, saliva, etc.).

ASSOCIATED CONTENT

S **Supporting Information**
Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: a.cifuentes@csic.es. Phone: +34-91-0017955. Fax: +34-91-0017905.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Gustavsson, A.; Brinck, P.; Bergvall, N.; Kolasa, K.; Wimo, A.; Winblad, B.; Jönsson, L. *Alzheimers Dement.* **2011**, *7*, 318−327.
(2) Blennow, K.; Hampel, H. *Lancet Neurol.* **2003**, *2*, 605−613.
(3) Ibach, B.; Binder, H.; Dragon, M.; Poljansky, S.; Haen, E.; Schmitz, E.; Koch, H.; Putzhammer, A.; Kluenemann, H.; Wieland, W.; Hajak, G. *Neurobiol. Aging* **2006**, *27*, 1202−1211.
(4) Cedazo-Minguez, A; Winblad, B. *Exp. Gerontol.* **2010**, *45*, 5−14.
(5) National Institute for Health and Clinical Excellence. *Dementia: Quick Reference Guide*, 42nd ed.; National Collaborating Centre for Mental Health: London, 2006.
(6) Hampel, H.; Prvulovic, D.; Teipel, S.; Jessen, F.; Luckhaus, C.; Froliche, L.; Riepe, M. W.; Dodel, R.; Leyhe, T.; Bertram, L.; Hoffmann, W.; Faltraco, F. *Prog. Neurobiol.* **2011**, *95*, 718−728.
(7) Johanson, C. E.; Duncan, J. A.; Klinge, P. M.; Brinker, T.; Stopa, E. G.; Silverberg, G. D. *Cerebrospinal Fluid Res.* **2008**, *14*, 5−10.
(8) Roche, S.; Gabelle, A.; Lehman, S. *Proteomics Clin. Appl.* **2008**, *2*, 428−436.
(9) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res.* **2009**, *37*, D603−D610.
(10) Czech, C.; Berndt, P.; Busch, K.; Schmitz, O.; Wiemer, J.; Most, V.; Hampel, H.; Kastler, J.; Senn, H. *PLoS One* **2012**, *7*, e31501.
(11) Myint, K. T.; Aoshima, K.; Tanaka, S.; Nakamura, T.; Oda, Y. *Anal. Chem.* **2009**, *81*, 1121−1129.
(12) Fonteh, A. N.; Harrington, R. J.; Tsai, A.; Liao, P; Harrington, M. G. *Amino Acids* **2007**, *32*, 213−224.
(13) Samakashvili, S.; Ibáñez, C.; Simó, C.; Gil-Bea, F. J.; Winblad, B.; Cedazo-Mínguez, A.; Cifuentes, A. *Electrophoresis* **2011**, *32*, 2757−2764.
(14) Sato, Y.; Suzuki, I.; Nakamura, T.; Bernier, F.; Aoshima, K.; Oda, Y. *J. Lipid Res.* **2102**, *53*, 567−576.
(15) Greenberg, N.; Grassano, A.; Thambisetty, M.; Lovestone, S.; Legido-Quigley, C. *Electrophoresis* **2009**, *30*, 1235−1239.
(16) Han, X. L.; Rozen, S.; Boyle, S. H.; Hellegers, C.; Cheng, H.; Burke, J. R.; Welsh-Bohmer, K. A.; Doraiswamy, P. M.; Kaddurah-Daouk, R. *PLoS One* **2011**, *6*, e21643.
(17) Simó, C.; Ibáñez, C.; Gómez-Martínez, A.; Ferragut, J. A.; Cifuentes, A. *Electrophoresis* **2011**, *32*, 1765−1777.

Analytical Chemistry

(18) Martín-Álvarez, P. J. In *Wine Chemistry and Biochemistry*; Moreno-Arribas, M. V., Polo, M. C., Eds.; Springer: New York, 2009; pp 677−714.

(19) Puerta, A.; Díez-Masa, J. C.; Martín-Álvarez, P. J.; Martín-Ventura, J. L.; Barbas, C.; Tuñón, J.; Egido, J.; de Frutos, M. *Analyst* **2011**, *136*, 816−822.

(20) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747−751.

(21) Peskind, E. R.; Li, G.; Shofer, J.; Quinn, J. F.; Kaye, J. A.; Clark, C. M.; Farlow, M. R.; DeCarli, C.; Raskind, M. A.; Schellenberg, G. D.; Lee, V. M.; Galasko, D. R. *Arch. Neurol.* **2006**, *63*, 936−939.

(22) Cedazo-Minguez, A. *J. Cell. Mol. Med.* **2007**, *11*, 1227−1238.

(23) Atkinson, W.; Hamborsky, J.; McIntyre, L.; Wolfe, S. *Epidemiology and Prevention of Vaccine-Preventable Diseases*, 11th ed.; Public Health Foundation: Washington, DC, 2009.

(24) Mattsson, N.; Rosén, E.; Hansson, O.; Andreasen, N.; Parnetti, L.; Jonsson, M.; Herukka, S. K.; van der Flier, W. M.; Blankenstein, M. A.; Ewers, M.; Rich, K.; Kaiser, E.; Verbeek, M. M.; Olde Rikkert, M.; Tsolaki, M.; Mulugeta, E.; Aarsland, D.; Visser, P. J.; Schröder, J.; Marcusson, J.; de Leon, M.; Hampel, H.; Scheltens, P.; Wallin, A.; Eriksdotter-Jönhagen, M.; Minthon, L.; Winblad, B.; Blennow, K.; Zetterberg, H. *Neurology* **2012**, *78*, 468−476.

(25) Gil-Bea, F. J.; Solas, M.; Solomon, A.; Mugueta, C.; Winblad, B.; Kivipelto, M. J.; Ramirez, M. J.; Cedazo-Mínguez, A. *J. Alzheimers Dis.* **2010**, *22*, 405−413.

(26) Wishart, D. S.; Lewis, M. J.; Morrissey, J. A.; Flegel, M. D.; Jeroncic, K.; Xiong, Y.; Cheng, D.; Eisner, R.; Gautam, B.; Tzur, D.; Sawhney, S.; Bamforth, F.; Greiner, R.; Li, L. *J. Chromatogr., B* **2008**, *871*, 164−173.

(27) Walter, A.; Korth, U.; Hilgert, M.; Hartmann, J.; Weichel, O.; Hilgert, M.; Fassbender, K.; Schmitt, A.; Klein, J. *Neurobiol. Aging* **2004**, *25*, 1299−1303.

(28) Arlt, S.; Schulze, F.; Eichenlaub, M.; Maas, R.; Lehmbeck, J. T.; Schwedhelm, E.; Jahn, H.; Böger, R. H. *Dement. Geriatr. Cogn. Disord.* **2008**, *26*, 58−64.

(29) Mulder, C.; de Jong, S.; van Kamp, G. J.; Scheltens, P.; Teerlink, T. *J. Neural Transm.* **2002**, *109*, 1203−1208.

(30) Rupsingh, R.; Borrie, M.; Smith, M.; Wells, J. L.; Bartha, R. *Neurobiol. Aging* **2011**, *32*, 802−810.

(31) Klopstock, T.; Elstner, M.; Bender, A. *Amino Acids* **2011**, *40*, 1297−1303.

(32) Foy, C. M. L.; Daly, E. M.; Glover, A.; O'Gorman, R.; Simmons, A.; Murphy, D. G. M.; Lovestone, S. *Brain Topogr.* **2011**, *24*, 316−322.

(33) Gualano, B.; Artioli, G. G.; Poortmans, J. R.; Lancha Junior, A. H. *Amino Acids* **2010**, *38*, 31−44.

(34) Flanagan, J. L.; Simmons, P. A.; Vehige, J.; Willcox, M. D.; Garrett, Q. *Nutr. Metab.* **2010**, *7*, 30−32.

S-1

SUPPORTING INFORMATION

Toward a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry Metabolomics

Clara Ibáñez[1], Carolina Simó[1], Pedro J. Martín-Álvarez[1], Miia Kivipelto [2],

Bengt Winblad[2], Angel Cedazo-Mínguez[2], Alejandro Cifuentes[1,*]

[1]Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain.

[2]Karolinska Institute, NVS Department, KI-Alzheimer's Disease Research Center,

14186 Stockholm, Sweden

Table of Contents

This supporting information file includes additional information as described in the text of the main article:

Experimental Section: Chemicals, Subjects, CSF sampling, CE-TOF MS analysis, Data processing, Statistical analysis.

Tables: S-1, S-2, S-3, S-4, S-5.

Figures: S-1, S-2, S-3.

S-2

2. EXPERIMENTAL SECTION

Chemicals

All chemicals were of analytical reagent grade, except reagents and solvents employed in the preparation of CE electrolytes and sheath liquid that were of MS grade, and all used as received. Namely, 2-propanol, methanol and water were from LabScan (Gliwice, Poland), acetic acid was from Scharlau (Barcelona, Spain) and formic acid was from Riedel-de Haën (Seelze, Germany). Tyramine and DL-methionine sulfone from Sigma-Aldrich (St. Louis, MO, USA) were selected as internal standards (IS) at final concentrations of 1 mg/L and 4 mg/L, respectively, in all the analyzed samples. Commercial standards from Sigma-Aldrich were dissolved in purified water deionized by using a Milli-Q system from Millipore (Bedford, MA, USA), at concentrations of 0.04 mg/mL for choline, arginine, histidine, carnitine, creatine, valine and serine; and 0.08 mg/mL for proline, dimethyl-arginine and leucine-proline.

Subjects

All patients underwent a complete clinical investigation according to the Karolinska University Hospital assessment protocol including medical inspection, brain imaging, electroencephalography, blood and CSF screening tests (including total tau, phospho-tau, and $A\beta_{1-42}$) and a detailed neuropsychological examination. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethics committee of the Karolinska University Hospital. Dementia and AD were diagnosed according to DSM-IV and NINCDS-ADRDA criteria. SCI subjects had cognitive complaints without impairment on objective cognitive tasks (subjective cognitive impairment), MCI patients were not demented and had (self and/or an informant) reported cognitive decline and impairment on objective cognitive tasks. MCI patients which progressed to AD within two-year time after the initial MCI diagnostic, preserved basic activities of daily living and minimal impairment in complex instrumental functions. Patients with psychiatric disorders (i.e. depression, alcohol abuse) or other conditions (i.e. diabetes, brain tumors, normal pressure hydrocephalus) at the time of CSF sampling were not included in this study.

CSF sampling

CSF extraction was routinely performed at the Karolinska University Hospital Memory Clinic in Huddinge (Sweden) as part of the medical examination. The extractions were performed in the mornings in fasting patients. Lumbar puncture was performed in the sitting position. CSF samples were obtained from L3/L4 or L4/L5 interspaces after local anesthetic infiltration in the

S-3

skin. 10 mL were collected in polypropylene tubes after disposal of the first milliliter. No sample contained more than 500 erythrocytes per microliter of CSF. Samples were gently mixed to avoid gradient effects and centrifuged at 2000×g for 10 min to eliminate cells and insoluble material. Supernatants were immediately aliquot and stored at -80 °C for pending biochemical and metabolomic analyses.

CE-TOF MS analysis

Before first use, the fused-silica separation capillary was conditioned by rinsing with 1 M NaOH for 10 min, followed by 20 min with Milli-Q water and 5 min with the separation buffer (0.5 M formic acid at pH 1.8). After each run, the capillary was conditioned with the separation buffer during 4 min. Electrical contact at the electrospray needle tip was established via a sheath liquid based on isopropanol-water (50:50, v/v) and delivered at a flow rate of 0.24 mL/h by a 74900-00-05 Cole Palmer syringe pump (Vernon Hills, IL, USA). The mass spectrometer operated with the ESI source in the positive ion mode. The nebulizer and drying gas flows were 0.4 bar N2 and 4 L/min N2 respectively, and maintaining the ESI chamber at 200ºC. Internal calibration of the electropherograms obtained by CE-MS was performed by injecting a 10 mM sodium formate solution for 4 minutes through the separation capillary at the end of each experiment. External calibration before each run was also carried out with 10 mM sodium formate solution. Masses to perform the calibration of the TOF MS instrument were: 90.9766, 158.9641, 226.9515, 294.9389, 362.9263, 430.9138 and 498.9012 m/z. Each sample was analyzed in duplicate by CE-MS.

Data processing

Peak detection was performed by successively using the chromatogram builder and peak deconvolution functions. For the chromatogram builder, the mass detection was achieved with the wavelet transform algorithm (noise level set at 1.50×103 a.u., scale level set at 6 a.u and a 50% wavelet window size) and the chromatogram construction was performed considering the highest data point with the following parameter values: minimum time span of 0:01, minimum height of 1×10^2 and m/z tolerance of 0.05. Then, for peak deconvolution, peak recognition was achieved using the "baseline cut-off" function (minimum peak height at 1×10^2, minimum peak duration at 0:01, baseline level at 1.50×10^3). Also a deisotoping step (m/z tolerance 0.01, RT tolerance 0:00, monotonic shape, maximum charge of 2) and an adduct search (show all possible adducts, max.adduct peak height with 100%) were applied. Next, a migration time normalization between duplicates was performed (m/z tolerance 0.01, time tolerance 0:10 and

S-4

minimum standard intensitiy of 3.50×10^4) before peak alignment using the "RANSAC aligner" (m/z tolerance at 0.01, RT tolerance after correction of 0:10, RT tolerance of 0:10, RANSAC iterations set at 0, minimum number of common points of 90%, threshold value at 0:30 by the linear model). Possible unspecific signals, noise and peaks not common between duplicates of the same sample were eliminated after a one-by-one manual revision from the peak list created by the program.

For sample alignment "Join aligner" was employed (m/z tolerance set at 0.03, weight for m/z at 90, relative retention time tolerance at 15%, absolute RT tolerance at 1:00, weight for RT at 10 and isotope pattern score threshold level of 50%).

Statistical analysis

To visually analyze processed data and conclude relevant remarks, a statistical workflow design and optimization is required, especially with convoluted data sets like the ones obtained from MS analysis of complex biological samples, and mainly when a profiling approach is applied. Among the most common statistical techniques in exploring and grouping MS data, principal component analysis (PCA) is a multivariate statistical unsupervised technique suitable for high dimensional data[1] as CE-TOF MS data and a tool often included in MS specific softwares. The main objective of this technique is to reduce the dimension of data without losing important information, to explore the relationship between variables (i.e. m/z values) and between samples, by creating new variables called principal components with the greatest variance in the data distribution. The bidimensional representation of the samples in the plane defined by the principal components is usually used to explore or confirm possible clusters of samples and to detect possible outliers.[2] Meanwhile, the main objective of multivariate statistical supervised classification techniques is to obtain classification rules from the information of the variables analyzed in the genuine samples (training data set), for *k* groups (i.e. diagnosis group). The resulting classification functions can be used to differentiate the defined *k* groups and to predict the classification of new samples (test data set) into one of the *k* groups.[3] The linear discriminant analysis (LDA) is one of the most frequently used supervised classification methods.[4] Similar results but selecting less variables are obtained by a forward stepwise LDA, that selects in each step, the variable that most differentiates the groups. Alternatively, a canonical variate analysis (CVA) is commonly employed to visualize results from the LDA classification method if *k*>2 obtaining a low-dimensional graphical representation of the samples,[5] where the *k* groups are separated as much as possible. To assess the predictive classification power of the classification method from the LDA, a leave-one-out cross-validation (LOO-CV) procedure is often used.[6]

S-5

The performance of a new diagnostic tool, in relation with an accepted one, is generally expressed as positive and negative predictive values (PPV and NPV), sensitivity and specificity. PPV can be defined as the ratio of true positives in relation to all the positive test results (true and false positives). The NPV is the ratio of true negatives in relation to all the negative test results (true and false negatives). Sensitivity can be defined as the ability of a diagnostic tool to detect the positive test results in relation to the total of true positives. A test with a high sensitivity has a low type II error rate or low number of false negative determinations. Specificity can be defined as the ability of a diagnostic tool to detect the negative test results in relation to the total of true negatives. A test with a high specificity has a low type I error rate or low number of false positive results. Satisfactory percentages (above 80%) for PPV, NPV, sensitivity and specificity were obtained in all binary LDA methods.

SUPPORTING INFORMATION REFERENCES

(1) Parra, L.; Deco, G.; Miesbach, S. *Neural Comput.* **1996**, *8*, 260–269.

(2) Hodge, V. J.; Austin, J. *Artif. Intell. Rev.* **2004**, *22*, 85-126.

(3) Martín-Álvarez, P. J. In *Wine Chemistry and Biochemistry*, Moreno-Arribas, M. V.; M.C. Polo Eds.; Springer: New York, 2009, 677-714.

(4) Toher, D.; Downey, G.; Murphy, T. B. *J. Chemometrics* **2011**, *25*, 621-630.

(5) Mitteroecker, P.; Bookstein, F. *Evol. Biol.* **2011**, *38*, 100-114.

(6) Subramanian, J.; Simon, R. *Statist. Med.* **2010**, *30*, 642-653.

S-6

Supporting Tables

Table S-1. 1.3. Age, sex and AD-related group of the 85 diagnosed subjects from whom the CSF samples (belonging to training set and test set in this work) were obtained.

Groups	nonAD		MCI-AD	AD	Total
	SCI-nonAD	MCI-nonAD			
Training set (n)	19	22	9	23	73
Age range (years)	45-65	43-77	51-73	55-84	
Sex (M/F)	7 M, 12 F	14 M, 8 F	4 M, 5 F	5 M, 18 F	
Test set (n)	4	2	4	2	12
Age range (years)	48-84	51-64	51-77	59-69	
Sex (M/F)	3 M, 1 F	2 M	3 M, 1 F	2 M	

S-7-9

Table S-2. Demographic characteristics and clinical diagnosis of studied subjects.

Patient ID	Group according disagnosis	Age (years)	Sex
3	SCI-nonAD	62	M
7	SCI-nonAD	53	M
11	SCI-nonAD	63	F
25	SCI-nonAD	49	F
30	SCI-nonAD	60	F
40	SCI-nonAD	61	M
42	SCI-nonAD	63	M
43	SCI-nonAD	51	M
45	SCI-nonAD	52	M
47	SCI-nonAD	56	F
49	SCI-nonAD	45	F
50	SCI-nonAD	59	F
51	SCI-nonAD	45	F
52	SCI-nonAD	64	F
54	SCI-nonAD	59	F
69	SCI-nonAD	62	F
70	SCI-nonAD	65	M
72	SCI-nonAD	61	F
74	SCI-nonAD	59	F
1	MCI-nonAD	64	M
12	MCI-nonAD	60	F
15	MCI-nonAD	62	F
31	MCI-nonAD	55	M
59	MCI-nonAD	57	M
64	MCI-nonAD	60	M
65	MCI-nonAD	73	M
67	MCI-nonAD	69	F
73	MCI-nonAD	49	F
77	MCI-nonAD	57	M
78	MCI-nonAD	77	M
79	MCI-nonAD	72	M
80	MCI-nonAD	43	F
86	MCI-nonAD	58	M
94	MCI-nonAD	51	M
102	MCI-nonAD	50	F
104	MCI-nonAD	67	M
105	MCI-nonAD	59	F
106	MCI-nonAD	65	M
108	MCI-nonAD	57	M
111	MCI-nonAD	42	F

Table S-2. (*Cont.*)

112	MCI-nonAD	59	M
61	MCI-AD	57	F
62	MCI-AD	60	M
82	MCI-AD	55	M
84	MCI-AD	73	M
85	MCI-AD	57	F
93	MCI-AD	57	F
101	MCI-AD	73	M
107	MCI-AD	71	F
110	MCI-AD	51	F
9	AD	79	F
13	AD	61	F
16	AD	67	F
17	AD	81	F
18	AD	78	F
19	AD	63	F
22	AD	55	F
23	AD	84	M
26	AD	63	F
37	AD	73	F
38	AD	81	F
39	AD	79	M
41	AD	55	F
44	AD	66	M
53	AD	78	F
55	AD	73	F
56	AD	60	F
63	AD	84	M
66	AD	55	F
75	AD	68	M
81	AD	63	F
88	AD	71	F
91	AD	74	F
2	Test (SCI-nonAD)	54	F
35	Test (SCI-nonAD)	84	M
96	Test (SCI-nonAD)	48	M
109	Test (SCI-nonAD)	68	M
29	Test (MCI-nonAD)	64	M
92	Test (MCI-nonAD)	51	M
97	Test (MCI-AD)	77	M
98	Test (MCI-AD)	51	F
103	Test (MCI-AD)	70	M
113	Test (MCI-AD)	65	M
8	Test (AD)	69	M
76	Test (AD)	59	M

S-10

Table S-3. Summary table on MZmine settings used in this work.

Chromatogram builder	Chromatogram decovolution	Deisotoping	Adduct search
Mass detection = Wavelet transform algoritm Noise level = 1.50×10^3 a.u Scale level= 6 a.u Wavelet window size= 50% Chromatogram construction = Highest data point: Min time span = 0:01 Min height = 1×10^2 m/z tolerance = 0.05	Peak recognition = Baseline cut-off Minimum peak height= 1×102 Minimum peak duration= 0:01 Baseline level= 1.50×103 m/z tolerance= 0.05	m/z tolerance = 0.01 Rt tolerance = 0:00 Mono tonic shape Maximum charge = 2 Representative isotope = Most intense	Show all possible adducts Max. adduct peak height= 100%
Migration time normalization (replicates)	**Alignment (replicates)**	**Alignment (samples)**	**Exporting**
m/z tolerance=0.01 time tolerance= 0:10 minimum standard intensity=3.50×104	RANSAC aligner: m/z tolerance = 0.01 Rt tolerance after correction= 0:10 Rt tolerance= 0:10 RANSAC iterations= 0 Minimum number of common points= 90% Threshold value= 0:30 by linear model	Join aligner: m/z tolerance= 0.03 Weight for m/z= 90 Relative Rt tolerance = 15% Absolute Rt tolerance= 1:00 Weight for Rt= 10 Isotope pattern score threshold = 50%	Export to CSV file Export MZMine ID, m/z, retention time, Peak area

Table S-4. Four-group comparison assignments by LDA and LOO-CV. Sample groups compared in each case are marked in the first column. Calculated sensitivity, specificity, positive and negative predictive value are also given for both classification procedures.

Group comparisons	Number of selected metabolites for the classification (in parenthesis ID of significant metabolites)	LDA					LOO-CV				
		% Correct classification	% Sensitivity	% Specificity	% Positive predictive value	% Negative predictive value	% Correct classification	% Sensitivity	% Specificity	% Positive predictive value	% Negative predictive value
SCI-nonAD vs. MCI-nonAD vs. MCI-AD vs. AD	**10** (7*, 3, 25, 65, 48, 43, 15, 1, 70, 2)	**90.1%** 94.7% (SCI-nonAD), 85.7% (MCI-nonAD), 88.9% (MCI-AD), 90.9%(AD)	-	-	-	-	**74.6%** 78.9%(SCI-nonAD), 71.4%(MCI-nonAD), 55.6%(MCI-AD), 81.8%(AD)	-	-	-	-
SCI-nonAD vs. MCI-nonAD	**8** (3, 2, 52, 4, 10, 58, 56, 17)	**95%** 95.2%(SCI-nonAD), 94.7%(MCI-nonAD)	95.2	94.7	95.2	94.7	**80%** 89.5%(SCI-nonAD), 71.4%(MCI-nonAD)	71.4	89.5	88.2	73.9
MCI-nonAD vs. MCI-AD	**3** (7, 53, 3)	**100%** 100%(MCI-nonAD), 100%(MCI-AD)	100	100	100	100	**96.7%** 95.2%(MCI-nonAD), 100%(MCI-AD)	100	95.2	90	100
MCI-AD vs. AD	**19** (60, 48*, 32, 7*, 11, 71*, 29* 12* 10*, 15, 50, 2, 41, 67, 53*, 59, 17, 1*, 38)	**100%** 100%(MCI-AD), 100%(AD)	100	100	100	100	**100%** 100%(MCI-AD), 100%(AD)	100	100	100	100
SCI-nonAD vs. MCI-AD	**20** (65, 25, 7, 3*, 60, 57, 66*, 40, 15, 56, 47, 59, 48, 41, 2*, 29, 55, 62, 17, 53)	**100%** 100%(SCI-nonAD), 100%(MCI-AD)	100	100	100	100	**100%** 100%(SCI-nonAD), 100%(MCI-AD)	100	100	100	100
SCI-nonAD vs. AD	**5** (25, 65, 12, 2, 30)	**97.5%** 100%(SCI-nonAD), 95.4%(MCI-AD)	100	94.7	95.5	100	**97.6%** 100%(SCI-nonAD), 95.5%(AD)	95.5	100	100	95
MCI-nonAD vs AD	**10** (48, 53, 7*, 19, 11, 36, 59, 15, 65, 31)	**100%** 100%(MCI-nonAD), 100%(AD)	100	100	100	100	**88.4%** 95.2%(MCI-nonAD), 81.8%(AD)	81.8	95.2	94.7	83.3

*More than one compound (adduct, complex...) corresponding to the same m/z value (the same metabolite)
Metabolite IDs order are given according to the forward stepwise LDA significance.

S-12-13

Table S-5. Identification of significant metabolites.

Metabolite ID	Migration time	Mean m/z	Groups classification	Formula	Tentative compound	Error (ppm)
1	5.56	203.148	*a, f, j*	$C_8H_{18}N_4O_2$	Symmetric dimethylarginine	-10,4
2	5.59	161.093	*a, d, f, g, h, j*	$C_6H_{12}N_2O_3$	Alanyl-alanine	8,9
3	5.65	170.079	*a, b, d, e, g, j*		Unknown	
4	5.67	106.991	*d*		Unknown	
7	5.89	104.107	*a, b, c, e, f, g, i, j*	$C_5H_{13}NO$	Choline*	2,6
10	5.97	133.096	*d, f*	$C_5H_{12}N_2O_2$	Ornithine	-8,8
11	6.00	147.111	*f, i*	$C_6H_{14}N_2O_2$	Lysine	-8,7
12	6.06	114.069	*f, h*	$C_4H_7N_3O$	Creatinine	2,0
15	6.24	175.119	*a, b, c, f, g, i*	$C_6H_{14}N_4O_2$	Arginine*	2,0
17	6.28	156.077	*b, d, f, g*	$C_6H_9N_3O_2$	Histidine*	4,4
19	6.39	120.101	*i*		Unknown	
21	6.43	203.147	*b*	$C_8H_{18}N_4O_2$	Dimethyl-L-arginine*	-13,7
25	6.85	162.112	*a, b, c, g, h, j*	$C_7H_{15}NO_3$	Carnitine*	-0,2
29	7.30	132.076	*b, f, g, j*	$C_4H_9N_3O_2$	Creatine*	-4,5
30	7.31	202.177	*h*		Unknown	
31	7.35	218.134	*c, i*		Unknown	
32	7.39	308.127	*b, c, f*	$C_{11}H_{21}N_3O_5S_1$	Tripeptide (G,T,M; V,S,C; M,S,A)	-0,3
35	7.54	229.154	*c, b*		Unknown	
36	7.58	246.166	*c, i*	$C_{12}H_{23}NO_4$	Valerylcarnitine	-14,2
38	7.65	90.055	*f*	$C_3H_7NO_2$	alanine, sarcosine	-3,4
40	7.73	298.050	*g*	$C_8H_{15}N_3O_5S_2$	L-Cysteinylglycine disulfide	-9,4
41	7.79	262.126	*f, g*	$C_{11}H_{19}NO_6$	Methylmalonylcarnitine	-8,1
43	7.92	198.081	*a, b, c*		Unknown	
44	7.95	212.0990	*j*	$C_{10}H_{13}NO_4$	Methyldopa, methyltyrosine	14,0
46	7.98	232.122	*b*	$C_{10}H_{17}NO_5$	Suberylglycine	13,7
47	8.37	233.111	*g*	$C_9H_{16}N_2O_5$	N2-Succinyl-L-ornithine/ 4-(Glutamylamino) butanoate	-9,6
48	8.09	118.087	*a, b, c, f, g, i,*	$C_5H_{11}NO_2$	Valine*	6,6
50	8.41	247.126	*f*	$C_{10}H_{18}N_2O_5$	L-gamma-glutamyl-L-valine,/ L-beta-aspartyl-L-leucine	-12,0
52	8.53	106.051	*b, d*	$C_3H_7NO_3$	Serine*	7,2
53	8.54	132.101	*c, e, f, g, i*	$C_6H_{13}NO_2$	Leucine	-4,8
55	8.65	283.138	*g*		Unknown	
56	8.90	120.066	*d, g*	$C_4H_9NO_3$	Threonine	8,6
57	8.93	133.060	*g*	$C_4H_8N_2O_3$	Asparagine	-3,3
58	8.99	150.058	*d*	$C_5H_{11}NO_2S$	Methionine	-2,6
59	9.05	147.076	*f, g, i*	$C_5H_{10}N_2O_3$	Glutamine	-2,7
60	9.10	315.128	*f, g*	$C_{14}H_{14}N_6O_3$	7,8-Dihydropteroic acid	5,9
62	9.22	176.101	*g*	$C_6H_{13}N_3O_3$	Citrulline	-12,4

Table S-5. (*Cont.*)

65	9.33	116.072	*a, c, g, h, i, j*	$C_5H_9NO_2$	Proline*	12,0
66	9.41	182.079	*g*	$C_9H_{11}NO_3$	Tyrosine	-13,7
67	9.42	123.055	*f*		Unknown	
70	9.99	144.103	*a*	$C_7H_{13}NO_2$	Proline betaine	6,9
71	10.27	136.044	*b, f*		Unknown	

*Metabolites confirmed by sample co-injection with commercial standard and subsequent CE-MS analysis.

Metabolites significant for the classification of:
a) 4 groups (SCI-nonAD *vs.* MCI-nonAD *vs.* MCI-AD *vs.* AD);
b) 3 groups (nonAD *vs.* MCI-AD *vs.* AD);
c) nonAD *vs.* AD;
d) SCI *vs.* MCI-nonAD;
e) MCI-nonAD *vs.* MCI-AD;
f) MCI-AD *vs.* AD;
g) SCI *vs.* MCI-AD;
h) SCI *vs.* AD;
i) MCI-nonAD *vs.* AD;
j) nonAD *vs.* MCI-AD.

S-14

Supporting Figures
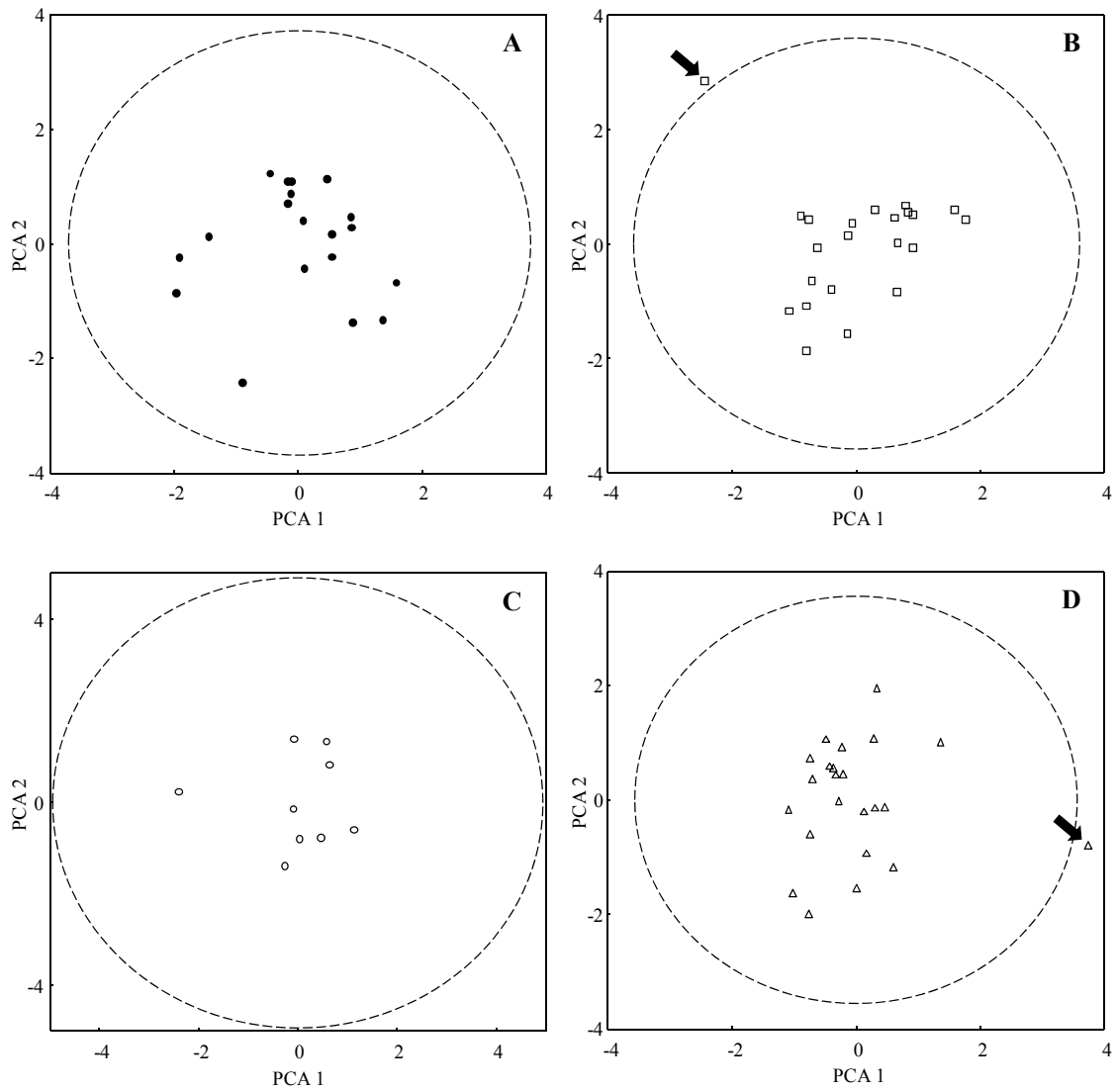
Figure S-1. Plot in a two-dimensional Cartesian coordinate system, with the axes (principal components) representing the greatest variations in the data of SCI-nonAD group (A), MCI-nonAD group (B), MCI-AD group (C) and AD group (D), is presented. 99% confidence ellipses are also included. Arrows indicate outliers.

S-15
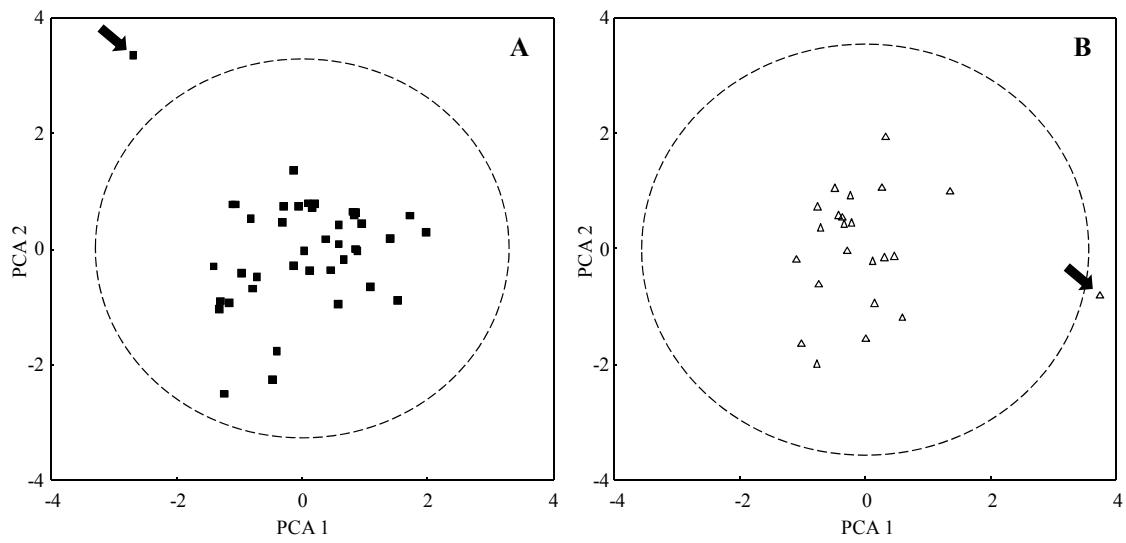
Figure S-2. Plot in a two-dimensional Cartesian coordinate system, with the axes (principal components) representing the greatest variations in the data of non-AD group (A) and AD group (B), is presented. 99% confidence ellipses are also included. Arrows indicate outliers.
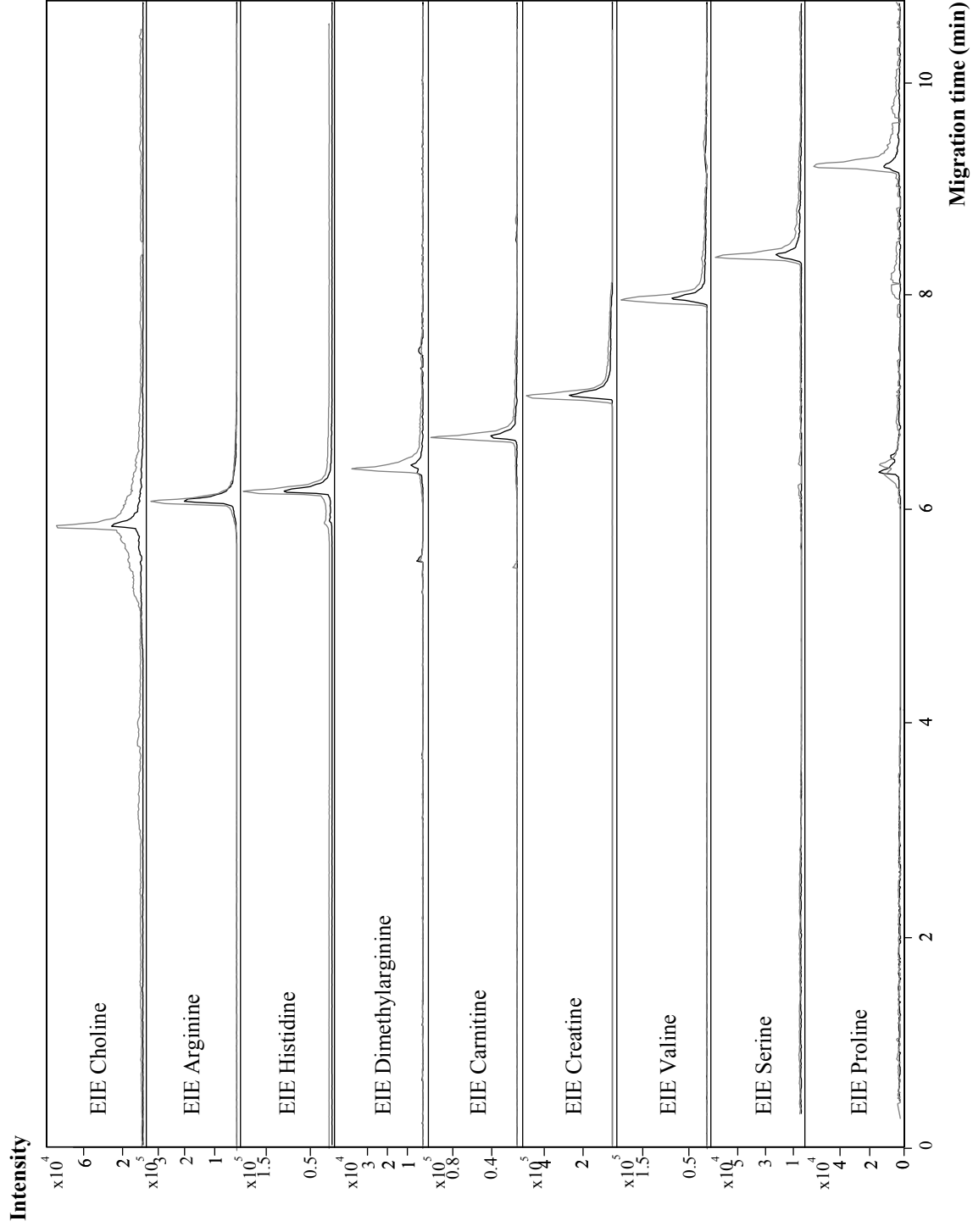
Figur e S-3. EIEs CSF (black line) and CSF+metabolite standards (grey line) coinjected. CE-MS conditions as in Figure 1.

4.5. CAPÍTULO 7 (Chapter 7)

A new metabolomic workflow for early detection of Alzheimer's disease

A NEW METABOLOMIC WORKFLOW FOR EARLY DETECTION OF ALZHEIMER'S DISEASE

Clara Ibáñez[a], Carolina Simó[a,*], Dinesh K. Barupal[b], Oliver Fiehn[b], Miia Kivipelto[c],

Angel Cedazo-Mínguez[c], Alejandro Cifuentes[a,*]

[a]Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain.

[b]University of California Davis, Genome Center, 451 E. Health Sci. Dr., Davis, CA 95616, USA.

[c]Karolinska Institute, NVS Department, KI-Alzheimer's Disease Research Center, 14186 Stockholm, Sweden.

Authors for the correspondance

*Address: Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain.

Carolina Simó:

e-mail: c.simo@csic.es. Telephone:+34-91-0017947. Fax: +34-91-0017905.

Alejandro Cifuentes:

e-mail: a.cifuentes@csic.es. Telephone: +34-91-0017955. Fax: +34-91-0017905.

ABSTRACT

Nowadays, Alzheimer's disease (AD), the most prevalent cause of dementia among older people, cannot be diagnosed accurately in early phases. However, AD probably starts 20-30 years before first clinical symptoms become noticeable making imperative the development of new methodologies for early diagnosis. In this work, we present a new metabolomics platform based on the use of ultra-high performance liquid chromatography-time-of-flight mass spectrometry (UPLC-TOF MS) to investigate cerebrospinal fluid (CSF) samples from patients with different AD stages. In order to obtain a metabolomics coverage as wide as possible, the developed metabolomics platform includes two different chromatographic separation modes, reverse phase (RP/UPLC-MS) and hydrophilic interaction chromatography (HILIC/UPLC-MS). RP/UPLC-MS and HILIC/UPLC-MS methods were optimized and applied to analyze CSF samples from 75 patients related to AD progression, investigating its potential as a tool for early diagnosis of AD. Significant metabolic differences in CSF samples from subjects with different cognitive status related to AD progression were detected by using this methodology, obtaining a group of potential biomarkers together with a classification model. The proposed model predicted the development of AD with an accuracy of 98.7 % and specificity and sensitivity values above of 95%.

KEYWORDS: Metabolomics; Alzheimer's disease; UPLC-MS; Multivariate statistical analysis.

1 INTRODUCTION

The incidence of many diseases increases rapidly with aging. Alzheimer's disease (AD), the most prevalent cause of dementia among older people, is a multifactorial disease in which age is the main risk factor. Although the initiating events are still unknown [1], AD results from a combination of genetic, environmental and lifestyle factors [2]. Nowadays, the scientific community faces important challenges related to the world's population aging and the projection indicating that the number of people aged 60 or older will grow from an estimated 600 million in 2010 to nearly 2 billion in 2050 [3]. In the "World Alzheimer Report 2009" [4] it was estimated that 35.6 million people would suffer dementia in 2010, the numbers nearly doubling every 20 years, to 65.7 million in 2030 and 115.4 million in 2050.

As public awareness of AD increases, the need for methodologies for early diagnosis is becoming imperative in an elderly population. However, early detection of AD is currently a huge challenge since AD probably starts 20-30 years before first clinical symptoms become noticeable. Moreover, early symptoms of AD are shared by various neuropathological disorders, including dementia. Up to date there is no clinical method to determine

which mild cognitive impairment (MCI) individuals will progress to AD except for a long clinical follow-up period. MCI condition is characterized for cognitive and memory problems that are not severe enough to be diagnosed as dementia but are more pronounced than the cognitive changes associated with normal aging. In approximately 80% of cases, MCI progresses to dementia when these subjects are followed up to 6 years [5].

Although there are several drugs that have been proven to slow disease progression and treat symptoms, so far, no treatment can effectively modify AD [6]. Early diagnosis of AD is expected to strongly impulse more research on potential AD risk factors, advances on more efficient drugs and cognitive stimulation programs.

Nowadays, AD can be definitively diagnosed only after death through an examination of brain tissue and pathology in an autopsy. However, it has been reported that increased total tau (t-tau) and phospho-tau (p-tau) while decreased amyloid beta (Aβ) levels have been observed in AD subjects in comparison with non-AD subjects.7 When these biomarkers are combined sensitivity and specificity values of 80-88% are obtained in the diagnosis of advanced cases [8,9]. On the other hand, advanced medical brain imaging techniques (computed tomography, nuclear magnetic resonance imaging and single photon or positron emission computed tomography) may help to diagnose the existence of dementia but not specific dementia due to AD [10]. The combination of the analysis of specific protein levels and brain imaging might increase AD diagnose up to 90% [9] increasing enormously the cost per patient (ca. 6,000-10,000 euros approximately). As imaging advances are being presented, parallel work is being carried out to identify reliable and valid markers in biofluids indicative of AD pathology [11]. In this sense, it can be observed that the search of biomarkers implicated in the pathogenesis of AD is moving away from traditional targeted to non-biased profiling approaches of human fluids in an attempt to discover novel biomarkers [12]. Among the "omic" platforms used to perform research into the biomarker discovery of AD, Proteomics has attained important consideration [7,13,14]. On the other hand, in the last decade Metabolomics has been demonstrated to be a promising discipline able to generate disease-specific metabolite signatures. It is concerned with the high-throughput identification and quantification of small molecules (<1500 Da). As for other age-related diseases [15], metabolomic study of biological fluids and tissues will also provide insights for a better understanding of molecular, structural, and functional changes that are causally related to the onset of AD. Although

Metabolomics is still in its infancy in AD investigation, several works have already focused on the study of lipid metabolites [16-19], since it has been hypothesized that lipid disfunctions can have some role in AD pathogenesis [20,21]. Recently, we have developed a new metabolomic approach based on capillary electrophoresis-mass spectrometry of CSF samples to investigate AD related to the ionic fraction of metabolites [22]. However, it is already well known that by using different metabolomics platforms complementary information can be obtained on other metabolites which usually provides a wider perspective of the problem under study [23]. In this sense, it is interesting to remark that at present the updated human CSF metabolome information has been obtained from five analytical platforms (NMR, GC-MS, LC-MS, DFI-MS/MS, ICP-MS) and contains 468 identified metabolites [24]. CSF is typically selected for metabolic profiling since it is the only body fluid in direct contact with the extracellular space of the brain and thus biochemical changes due to pathological brain-processes are more probable to be reflected in CSF than in other body fluids.

The first goal of this work is to develop an ultra-high performance liquid chromatography-time-of-flight-mass spectrometry (UPLC-TOF MS) approach for the non-targeted analysis of mid/high polarity metabolites in CSF without previous derivatization. Two complementary stationary phases for metabolite separation are used to cover a broader polarity range of metabolites and increase metabolome information. As stated above MCI has been regarded as a pre-Alzheimer condition, but some patients after a follow-up period do not develop dementia. Thus, the second goal of this work is to improve both sensitivity and specificity in the diagnosis of early AD patients using a metabolomic approach based on UPLC-TOF MS and multivariate statistical analysis of CSF samples. For this purpose, subjects enrolled in the study were classified into four diagnostic groups: healthy controls, MCI subjects who developed AD after a follow-up period of 2 years, MCI subjects whose initial MCI status remained stable during the course of the follow-up, and a group of subjects with AD. Metabolic differences in CSF samples from subjects with different cognitive status related to AD progression will be discussed in this work.

2 EXPERIMENTAL SECTION

2.1 Subjects

In this study we recruited 75 participants who underwent a complete investigation. Detailed ethics statements and methods used for clinical diagnosis of patients are included in appendix A. Briefly, among all the individuals of the study, 21

subjects (control group) did not present dementia and history of central nervous system (CNS). Among the 33 individuals who presented mild cognitive impairment (MCI) at baseline, after a follow-up period of 2 years, 12 developed AD (MCI-AD group) and 21 remained stable (MCI-S group). Lastly, 21 patients were diagnosed with mild AD (AD group). Demographic and clinical information of the four groups of the study is presented in Table 1.

2.2 CSF sampling and metabolite extraction

CSF sampling was routinely performed at the Karolinska University Hospital Memory Clinic in Huddinge (Sweden) as part of the medical examination (detailed description is given as supporting information in Appendix A). Metabolite extraction was performed as described elsewhere [22]. Briefly, prior to metabolite extraction, CSF samples were spiked with internal standards (tyramine and methionine sulfone at final concentration of 1 mg/L and 4 mg/L, respectively). The metabolite extraction procedure was as follows: 100 µL of CSF with added IS was ultrafiltrated by using 3 kDa Amicon Ultra 0.5 mL centrifugal devices from Millipore (Billerica, MA, USA). The centrifugation was performed at 14000×g for 40 min at 20° C. Filtered fractions (compounds with less than 3 kDa) were directly analyzed by UPLC-TOF MS analysis.

2.3 Chemicals

Water and organic solvents were of MS grade. Metabolite standards were of analytical grade and used as received. Detailed description of chemicals is given as supporting information in Appendix A.

2.4 Metabolomic analyses

Metabolomic analyses were performed using a UPLC system 1290 connected to a Q/TOF MS 6540 from Agilent (Agilent Technologies, Santa

Clara, CA, USA) equipped with an orthogonal ESI interface (Agilent Jet Stream, AJS) and operating in positive ion mode. The instrument was controlled by a PC running the Mass Hunter Workstation software 4.0 (MH) from Agilent. MS operation parameters were the following: capillary voltage, -4000 V; nebulizer pressure, 30 psi; drying gas flow rate, 10 L/min; gas temperature, 300 °C; skimmer voltage, 45 V; fragmentor voltage was 125 V in positive mode. TOF MS accurate mass spectra were recorded across the range of 50-1000 m/z at 1.5 spectra/s. Further specifications on internal and external calibration of the instrument are described in Appendix A.

Reverse-phase (RP) chromatographic separation was performed on an Agilent ZORBAX C18 Rapid Resolution HD column (2.1 × 50 mm, 1.8 µm) maintained at 40 ºC. Elution was performed using as phase A (water with 0.1% formic acid), and as phase B (acetonitrile with 0.1% formic acid), and the following gradient program: 0% B in 0-1 min, 0-2% B in 1-4 min, 2-20%B in 4-6 min, 20-100% in 6-10 min, 100% B in 10-12 min. After analysis, the column was re-equilibrated for 4 min using the initial solvent composition. Hydrophilic interaction chromatography (HILIC) was performed on a ZORBAX HILIC Plus HD (2.1 × 50 mm, 1.8 µm) column maintained at 30ºC. Elution was performed using a phase A (10 mM ammonium formate at pH 5.8 in water) and a phase B (acetonitrile) and the following gradient program: 95-90% B in 0-3 min, 90-80% B in 3-7 min, 80-0%B in 7-9 min, 0%B in 9-10 min. Then 95% B was kept constant for 6 min for column re-equilibration. Further information on quality control, blanks and replicates for both chromatographic methods is given in Appendix A.

2.5 Data processing

Raw UPLC-MS centroid data were converted to the MS exchange format mzXML using the open-

Table 1. Clinical information of the participants of the study.

Groups	**Control**	**MCI-S**	**MCI-AD**	**AD**	Total
Participants (n)	21	21	12	21	75
Age average ± S.D. (age range)	58 ± 8.9 (45-84)	60 ± 8.9 (42-77)	63 ± 9.4 (51-77)	69 ± 9.6 (55-84)	62 ± 9.9 (42-84)
Gender M/F (M%/F%)	9/12 (43/57)	14/7 (67/33)	6/6 (50/50)	6/15 (29/71)	35/40 (47/53)

source program Trapper version 4.3.0 (available at http://tools.proteomecenter.org/wiki/index.php?title=Software:trapper). mzXML files were pre-processed with XCMS, an open-source package [25] written in the platform-independent programming language R (www.r-project.com). Peak detection method used in XCMS package [25] was the CentWave algorithm [26]. The resulting sample files were then converted to peakML format to execute the mzMatch.R package [27] in R software. The package mzMatch.R groups the features and performs noise filtering, gap filling and related peaks annotation. The resulting .txt file was then imported into IDEOM version 14 [28] to automatically check sample consistency in terms of standard deviation of the internal standards and to examine and revise the detected features. Filtering was then carried out to ensure ions with a high quality: (i) peaks below 3x intensity from the blanks were removed, (ii) peaks not found in at least a 75% of the samples of the same classification group or with a high variability within the same group (with a value of median/average > 1.5) were removed. The resulting output data table of high quality time-aligned detected peaks (also called "variables" along the text), with their corresponding migration time, molecular weight and peak area obtained for each sample, was used to perform a multivariate statistical analysis. Additional details regarding XCMS, mzMatch and IDEOM are summarized in Table A-1.

2.6 Multivariate statistical analysis

First a Principal Component Analysis (PCA) was applied to detect possible outliers in each group of samples (Control, MCI-S, MCI-AD and AD groups). Once outliers were discarded, a Levene's test was applied to assess the equality of the variances (similar standard deviations) of the variables (i.e. metabolites) in all the groups. Analysis of variance by a parametric one-way ANOVA was performed to detect the different peak areas averages (p<0.05). A non-parametric analysis of variance applying a Kruskal-Wallis test was performed to compare the medians of the variables with heterogeneity in the variances among the four groups (revealed by Levene's test). Linear Discriminant Analysis (LDA), with forward stepwise procedure (with probabilities values set at 0.05 to enter and to remove variables) was performed to select the significant variables (p<0.05) most useful in differentiating the four sets of samples. Canonical Variate Analysis (CVA) was also carried out to obtain a low-dimensional graphical representation of the samples that separates the groups as much as possible. In addition, three different "multiple comparison procedures" were applied namely LSD, Scheffe and Games-Howell tests. These tests were

performed to determine the significance level of the metabolites selected by the LDA for all two-group comparisons. Whisker plot were used to graphically examine the differences in the average values corresponding to the LDA selected variables. The predictive power of the LDA classification method was carried out by using a Leave-One-Out Cross-Validation (LOO-CV) procedure [29,30]. In clinics, the use of diagnostic test indicators to compare a new methodology to a well-established one is done very often. For this reason, binary LDA classification models for all the possible two-by-two group comparisons was performed and positive and negative predictive values (PPV and NPV), sensitivity and specificity were calculated as diagnosis test indicators (further description is given as supporting information in Appendix A). STATISTICA (v.9, Statsoft, Tulsa, OK, USA, www.statsoft.com) and SPSS (v.19, IBM, Chicago, IL, USA, www.spss.com) programs for Windows were used for the statistical analysis.

2.7 Metabolite identification and related pathways

Significant metabolic features were identified by means of IDEOM by matching the accurate mass of observed peaks to metabolites in the selected databases (KEGG [31], HMDB [32], Metacyc [33], and LipisMaps [34]) within a mass accuracy window of 10 ppm. Isotopic pattern was also considered to support the putative identification. Metabolic pathways associated to metabolic classifiers were studied. Thus, all the metabolites highlighted by the LDA were mapped into KEGG pathways to elucidate the metabolic processes more affected by the progression of the disease. Additional details regarding metabolite identification are given in the supporting information at Appendix A.

3 RESULTS AND DISCUSSION

3.1 Analysis of CSF samples by UPLC-TOF MS

Metabolic extracts from CSF samples under study were analyzed using the two analytical platforms described in section 2.4. Relative standard deviations for quality control injections were less than 1% and 11% for retention time and peak areas respectively. Differences observed in metabolic profiles from RP/UPLC-TOF MS and HILIC/UPLC-TOF MS were assigned to their different separation mechanisms. After peak detection, alignment and grouping using both XCMS and mzMatch.R packages in R program, 3546 and 11159 features (defined as a unique m/z value at a unique time point) were obtained from HILIC/UPLC-MS and RP/UPLC-MS analysis, respectively. After filtering process with IDEOM to study only the ions with a high quality present in

all the samples, 286 peaks were obtained from HILIC/UPLC-MS analysis and 238 peaks were obtained from RP/UPLC-MS analysis. This final matrix containing 524 high confident time-aligned peaks was subjected to multivariate statistical analysis.

3.2 Multivariate modeling

PCA was firstly applied to each group of samples in order to detect possible outliers before the application of any supervised statistical analysis. One sample was considered outlier (sample 34 from control group) as it was placed out of the 95% confidence ellipse in the PCA graphical representation (see Figure A-1), and thus, excluded for the subsequent statistical analysis. Analysis of variance was performed by means of both, a parametric method (ANOVA) and a non-parametric test (Kruskal-Wallis) at a significance level of 0.05. The one-way ANOVA was applied to all the set of variables to compare the mean levels of the metabolites among groups. As a result, 145 variables (i.e. metabolites) were differently expressed (p<0.05). From these 145 variables, 76 variables with heterogeneity on their variances (revealed by Levene's test) were submitted to the Kruskal-Wallis test that evaluates the variance of the medians among groups. 12 variables out of 76 were not significantly different (p<0.05) and discarded for the subsequent statistical analysis. The application of these two methods for the analysis of the variance prevents the consideration of false significant differences due to chance. The set of 133 variables significantly different by both, ANOVA and Kruskall-Wallis, were then used to build a classification model by a forward stepwise LDA (with probabilities values set at 0.05 to add and to remove variables). After this analysis 17 peaks with homogeneity in their variance among groups and normal distribution were selected to build the classification model. When applying the classification model, all the samples were assigned to the correct classification group except one sample (sample 75) from MCI-AD group which was misclassified as control group. On the other hand, when the classification model was applied to classify the outlier sample from the control group (sample 34), it was correctly assigned as control group. Therefore, a 98.7% of samples were properly assigned to their diagnostic group (Table A-2). Moreover, all possible binary LDA classification models for the two-by-two group comparisons were carried out to determine the metabolites responsible of each AD stage development. Satisfactory values above 95% of correct assignments of the samples were obtained for all the LDA classification models (see Table A-3). The lowest percentage of correct classification (95.2%) was obtained when control and MCI-S

groups were compared. One out of 21 control samples (sample 29) was misclassified as MCI-S, and one out of 21 MCI-S samples (sample 17) was misclassified as control group. The same value of 95.2% was obtained for sensitivity, PPV and PNV. A higher correct classification percentage of 97.6% was obtained when MCI-S and AD were compared (sample 17 was misclassified in AD group). 100% of sensitivity, 95.2% of specificity, 95.2% of PPV and 100% of NPV were calculated in this occasion. A correct classification percentage of 97.6% was obtained when control and AD groups were compared (sample 21 from AD group was misclassified as control group) applying the LDA model. In this case a 95.2% of sensitivity, 100% of specificity, 100% of PPV and a 95.5% of NPV were achieved. Finally, for all the other binary LDA classification models a 100% of correct classification as well as 100% in the four diagnostic test indicators were achieved. Considering the 17 selected variables (metabolites that most differentiated the four groups of samples), a CVA was applied to the 75 samples, and plotted in the space defined by the three calculated canonical variables (Figure 1). In order to assess the predictive power of the LDA method with the 17 selected used to build the classification model, LOO-CV was applied. LOO-CV classification was carried out by holding each sample out (one-at-a-time) and building a LDA model from the remaining samples and then classifying the held out sample using this model. A total LOO-CV satisfactory value of correct assignment above 85% was obtained proving the high accuracy prediction of the LDA model (see Table A-4). To deeply study differences among the levels of these 17 metabolites (shown in Table A-5), Scheffe, LSD, and Games-Howell tests were performed. These tests displayed the p-values assigned for all the two-group comparisons showing for all the metabolites at least one significant difference at p<0.05 (see Table A-6). These "multiple comparison procedures" use different statistical procedures to determine the significance level of the variables under study. The simultaneous application of these three methods provided a wide view of the statistical variables behavior and high confident information.

3.3 Potential biomarkers of AD progression

In order to study the changes in the metabolism due to AD progression, a more exhaustive identification focused on the significant peaks selected in all the forward stepwise LDA methods was carried out. Tentative identification of the significant variables obtained from the LDA analysis is given in Table A-7 (ID from 1 to 17). In addition, some metabolites that were exclusively selected in some of the binary LDA classification models were also identified (ID from 18 to 24 in
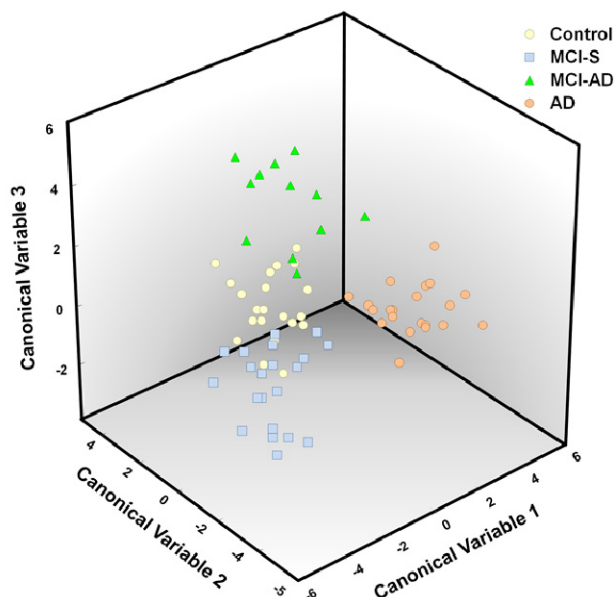
Figure 1. 3D-CVA plot of the 75 samples in the plane defined by the first three canonical variables. Representation obtained considering the 17 metabolites selected by the forward stepwise LDA. Control, MCI-S, MCI-AD and AD subjects are represented by yellow circles, blue squares, green triangles and orange circles, respectively.

Table A-7). Tentative identification of the 17 selected variables which most differentiated the four groups of samples according to the forward stepwise LDA were: uracil, xanthine, uridine, tyrosyl-serine (or serinyl-tyrosine), methylsalsolinol, nonanoylglycine, dopamine-quinone, caproic acid, vanylglycol, histidine, pipecolic acid, hydroxyphophinyl-piruvate, creatinine, taurine, C16-sphingosine-1-phosphate, tryptophan and 5'-methylthioadenosine. In Figure 2 related whisker plots for these metabolites are shown. From this group, uracil, histidine, creatinine, taurine, tryptophan and methylthioadenosine, were confirmed by coinjection of commercial standards with the sample..

3.4 Biological meaning

Metabolites uracil, taurine, nonanolglycine, caproic acid and sphingosine-phosphate continuously increased from the Control to the MCI-AD group and they suddenly decreased in AD patients. An opposite behavior was observed for tryptophan and pipecolic acid that continuously decreased from the Control to the MCI-AD group and increased in AD patients. Creatinine showed a continuous decrease from Control to AD patients, while uridine and dopamine-quinone showed the opposite behavior, a continuous increase from Control to AD patients. Histidine and tyrosil-serine showed a minimum concentration for MCI-AD patients, while xanthine, sphingosine-1-phospate

and hydroxyphosphinyl-piruvate presented a maximum concentration in MCI-AD patients. It is interesting to remark the relation with AD already described in literature for many of these possible biomarkers of Figure 2 while some of them are for the first time related to AD in this work as discussed next.

Sphingosine-1-phosphate, a lipid mediator generated from the sphingosine kinase-catalyzed phosphorylation of sphingosine, is now recognized as a critical regulator of many kinds of physiological and pathological processes [35]. More precisely, sphingosine-1-phosphate has been identified in neural cells to be involved in differentiation, growth, cell migration, and apoptosis and its abnormal metabolism may be closely associated with pathogenesis of Alzheimer's disease [36]. Uridine is the major form of pyrimidine nucleosides taken up by the brain and phosphorylated to nucleotides, which are used for DNA and RNA synthesis. Furthermore, it has been described that uridine administration has sleep-promoting and anti-epileptic actions, improves memory function and affects neuronal plasticity. Although, the exact role of pyrimidine nucleosides in neuronal functions is not established yet, they have been suggested for the treatment of epileptic and neurodegenerative diseases as neuroprotective agents. In addition, the development of traditional drugs acting specifically on pyrimidine receptor subtypes is also a promising direction to treat neurological

190
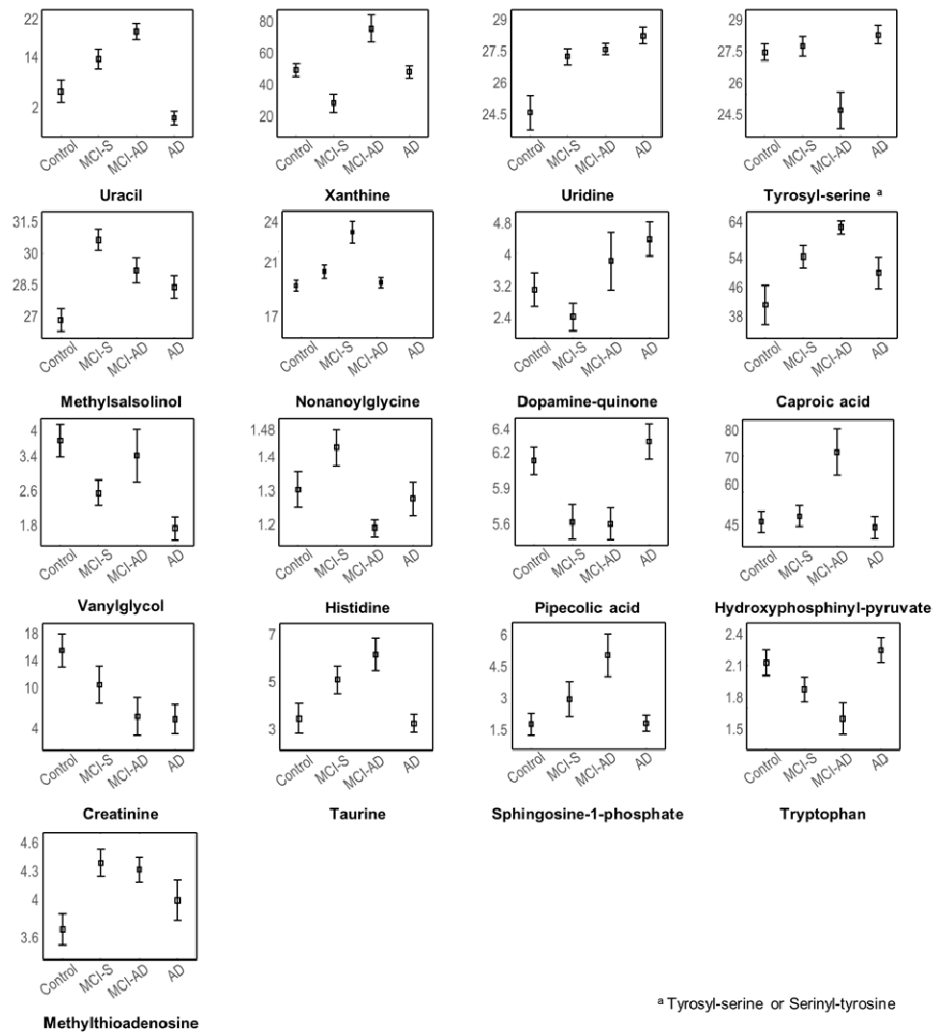
Figure 2. Whisker plot representing the selected metabolites mean values ($x10^3$) with the standard error bars associated. The standard error (SE) was calculated according to the next formula: SE=SD√n; where SD is the standard deviation and n is the number of observations.

disorders [37]. However, our results seem to indicate that uridine is not correctly metabolized by AD patients what could explain the higher levels of this compound observed for this group of patients. In our study, taurine, an inhibitory neurotransmitter, was observed to decrease in AD patients (see Figure 2). This inhibitory neurotransmitter has been demonstrated to have neuroprotective and neurotrophic properties helping to the growth and survival of developing neurons and the maintenance of mature neurons, the global lower brain functionality in a neurodegenerative disease as AD could explain the observed decrease of taurine in AD patients [38]. Tryptophan has extensively been related to AD. Using transgenic mice as model of AD it has been demonstrated that a diet rich in tryptophan brings about an increase in serotonergic (5-HT) neurotransmission suggesting that it may be effective in reducing plaque pathology in AD [39].

Besides, several studies have already been carried out to assess the efficacy of a docosahexenoic acid (DHA)-phospholipids, melatonin and tryptophan supplemented diet in improving the erythrocyte oxidative stress, membrane fluidity and membrane-bound enzyme activities of elderly subjects suffering from mild cognitive impairment [40]. It has been also mentioned that different neurodegenerative disorders, e.g. Parkinson's, Huntington's and AD, although being distinct clinical and pathological entities, they share a number of leading features in their underlying processes. These common features involve the disturbances in the normal functioning of the mitochondria and the alterations in the delicate balance of tryptophan metabolism [41].
Purine metabolites (creatinine and xanthine) were investigated in CSF of AD patients years ago correlating their content with the energy state of the brain [42]. Our results are in agreement with

the observed results about the increase of xanthine in AD patients, however, a decreasing concentration of creatinine in CSF of AD patients was observed here while in the mentioned work the authors observed that creatinine increased [42]. In this regard, it has been recently mentioned that some other important factors have to be considered to correlate purine metabolites detected in CSF to AD, including the renal function and the integrity of the blood–brain barrier [43].

Similar differing results were observed for uridine, a nucleoside that has been suggested to play a role in neurodegenerative diseases. In our study, uridine was observed to increase in CSF of AD patients (see Figure 2). However, in a recent work, uridine was observed to decrease in the CSF of AD patients, what according to the authors, could reflect reduced synaptic plasticity and neuronal deficits [44]. These contradictory results can be explained considering that the specific effects of nucleosides depend also on the expression of their receptors and transporters in neuronal and glial cells, as well as their extracellular concentrations in the brain. In addition, nucleoside levels depend also on brain region, age and gender, suggesting different roles of nucleosides in functionally different brain areas [45]. Besides, uridine is one of the four basic components of RNA and upon digestion of foods containing RNA, uridine is released from RNA and is absorbed intact in the gut. Therefore, an important effect of diet on these results cannot be ruled out either.

Dopamine-quinone, an oxidized product of dopamine, was observed in our work to increase in CSF of MCI-AD and AD patients (see Figure 2). In agreement with our finding, this metabolite has been linked to AD due to its neurotoxic activity, namely, dopamine-quinone which facilitates the aberrant aggregation of microtubule associated protein tau and participate in tau polymerization, which is one of the main characteristic of neurodegenerative disorders including AD [46].

Vanylglycol (also called 4-hydroxy-3-methoxyphenylglycol, HMPG) is a methylated metabolite of normetanephrine. HMPG is formed by catechol-O-methyltransferase from norepinephrine. This monoamine metabolite showed a large variability in the CSF samples from the different groups (Control, MCI-S, MCI-AD and AD) as can be seen in Figure 2. These results are in agreement with two previous works in which HMPG was determined in CSF of 123 patients with AD and 57 healthy controls. A wide variability in values among both patients and controls was found, as well as fluctuations in repeated samples from individual patients [47]. Variability for this monoamine metabolite was also corroborated in another work in which HMPG was determined in samples linked to frontotemporal dementia and AD [48].

Methylsalsolinol, an endogenous neurotoxin in the brain, is one of key causative agents of the death of dopaminergic neurons. There is increasing evidence that this type of endogenous production of toxins may be involved in the development of a number of neurodegenerative diseases such as Alzheimer's, Parkinson's or Huntington's disease, indicating that the mechanisms leading to cell loss are a combination of oxidative stress, mitochondrial dysfunction and a decrease in antioxidant defenses [49]. The concentration of methylsalsolinol was found to increase significantly in the CSF of patients with Parkinson's disease [50]. However, we have not found any previous correlation of the levels of methylsalsolinol in CSF with AD progression; moreover, the role of salsolinol N-methyltranseferase, a key enzyme to metabolize salsolinol into methylsalsolinol has to be also considered in the pathogenesis of AD. These data could partially explain the variability observed of methylsalsolinol in the four groups under study (see Figure 2), however, more work is necessary to clarify the role of this specific compound in AD progression.

In this work, both uracil and uridine come out as good candidate biomarkers. While it has been observed that uracil decreases, uridine increases in AD patients with respect to control subjects (see Figure 2). This fact could be associated with an alteration on the reversible conversion of uridine to uracil by the enzyme uridine phosphorilase. This enzyme has been suggested to present functions in sensing and initiating cellular responses to oxidative stress [51], process previously associated with a loss of neurons [49]. On the other hand, uracil is the main substrate of another key regulatory enzyme of oxidative stress in the brain (uracil-DNA glycosylase). This enzyme removes oxidized pyrimidines and is overexpressed under oxidative stress as a mechanism of protection from neurodegeneration [52]. Decreased levels of uracil in AD patients observed in this work could result in an alteration of this protection leading to neurodegeneration. Furthermore, uracil acts via channel P2X in the brain. This channel controls neurotransmission, neuromodulation, cell proliferation, differentiation and death, and their alteration has been related to several neurodegenerative diseases including AD [53]. The blockade of these channels has been associated with a reduction on proinflammatory mediators, enhancing neuroprotection by reducing the amyloid plaque formation, via non-amyloidogenic process [54]. With all this information and considering that uracil is the most important metabolite in classifying among all the diagnostic groups (first selected in the LDA), it is evident that this pyrimidine nucleobase should be deeply studied in order to determine more about its

implication in AD progression.

This work also shows for the first time a possible relation between different metabolites as caproic acid (hexanoic acid), nonanoylglycine, tyrosil-serine (or serinyl-tyrosine), hydroxyphosphinyl-pyruvate, pipecolic acid, methylthioadenosine and AD. For instance, pipecolic acid (piperidine-2-carboxylic acid) is a small organic molecule which accumulates in pipecolic acidemia and it has been associated with some forms of epilepsy [55], however, it has never been linked to AD progression before. Likewise, methylthioadenosine (MTA) is a naturally occurring sulfur-containing nucleoside present in all mammalian tissues. MTA is produced from S-adenosylmethionine mainly through the polyamine biosynthetic pathway, where it behaves as a powerful inhibitory product. Abundant evidence has accumulated over time suggesting that MTA can influence numerous critical responses of the cell including regulation of gene expression, proliferation, differentiation and apoptosis. Observations carried out in models of liver damage and cancer demonstrate a therapeutic potential for MTA that deserves further consideration [56].

4 CONCLUDING REMARKS

AD probably starts 20-30 years before first clinical symptoms become noticeable. In this work, we have developed a new metabolomics approach based on the UPLC-MS analysis of CSF samples related to AD progression in order to investigate its potential as a tool for early diagnosis of AD. The proposed model predicted the development of AD with an accuracy of 98.7 % and specificity and sensitivity values above of 95%. Moreover, a group of potential biomarkers has been identified in a first discovery phase. Nevertheless, a second validation phase should be followed on a large number of samples that closely reflect the population where the clinical test would be applied. Since urine and blood-based biomarkers offer a minimally invasive method for the diagnosis of late-life diseases, further research is urgently needed in this field. This kind of biofluid would accelerate and reduce the cost of AD diagnosis and monitoring. Future studies in our research group will also be focused on developing predictive tools able to correlate CSF changes with blood variations in AD progression.

Conflict of interest statement
The authors declare no conflict of interest.

ACKNOWLEDGMENTS

REFERENCES
[1] C.L. Masters, K. Beyreuther, Brain 129 (2006) 2823.
[2] X.F.Meng, C.D'Arcy, PLOS ONE 7 (2012) e38268.
[3] J.Chamie, World Population Ageing 1950-2050, United Nations Publications, New York, 2002.
[4] M. Prince, J. Jackson, World Alzheimer Report 2009, United Nations Publications, United Kingdom, 2009.
[5] R.C. Petersen, J. Intern. Med. 256 (2004) 183.
[6] A.Corbett, C. Ballard, Expert. Opin. Emerg. Dr. 17 (2012) 147.
[7] K.Blennow, H. Hampel, Lancet. Neurol. 2 (2003) 605.
[8] B.Ibach, H.Binder, M.Dragon, S. Poljansky, E. Haen, E. Schmitz, H. Koch, A. Putzhammer, H. Kluenemann, W. Wieland, G. Hajak, Neurobiol. Aging (2006) 27 1202.
[9] A. Cedazo-Minguez, B. Winblad, Exp. Gerontol. 45 (2010) 5.
[10] A. Fairbairn, N. Gould, T. Kendall, Dementia: supporting people with dementia and their careers in health and social care. Quick reference guide, National Institute for Health and Clinical Excellence, London, 42nd ed., 2006.
[11] R. Craig-Schapiro, A.M. Fagan, D.M. Holtzman, Neurobiol. Dis. 35 (2009) 128.
[12] R. Ghidoni, L. Benussi, A. Paterlini, V. Albertini, G. Binetti, E. Emanuele, Neurodegener. Dis. 8 (2011) 413.
[13] S. Lista, F. Faltraco, H. Hampel, Progr. Neurobiol., in press.
[14] M. Thambisetty, S. Lovestone, Biomark. Med. 4 (2010) 65.
[15] R.J. Mishur, S.L. Rea, Mass Spectrom. Rev. 31 (2012) 70.
[16] X. Han, S. Rozen, S.H. Boyle, C. Hellegers, H. Cheng, J.R. Burke, K.A. Welsh-Bohmer, P.M. Doraiswamy, R. Kaddurah-Daouk, PLOS ONE 6 (2011) e21643.
[17] M. Orešič, T. Hyötyläinen, S. K. Herukka, M. Sysi-Aho, I. Mattila, T. Seppänan-Laakso, V. Julkunen, P.V. Gopalacharyulu, M. Hallikainen, J. Koikkalainen, M. Kivipelto, S. Helisalmi, J. Lötjönen, H. Soininen, Transl. Psychiatry 1 (2011) e57.
[18] R. González-Domínguez, T. García-Barrera, J.L. Gómez-Ariza, Chem. Pap. 66 (2012) 829.
[19] Y. Sato, I. Suzuki, T. Nakamura, F. Bernier, K. Aoshima, Y. Oda, Lipid Res. 53 (2012) 567.

[20] G. van Echten-Deckert, J. Walter, Progr. Lipid Res. 51 (2012) 378.

[21] C.R. Hooijmans, A. Kiliaan, J. Eur. J. Pharmacol. 585 (2009) 176.

[22] C. Ibáñez, C. Simó, P.J. Martín-Álvarez, M. Kivipelto, B. Winblad, A. Cedazo-Mínguez, A. Cifuentes, Anal. Chem. 84 (2012) 8532.

[23] C. Ibáñez, C. Simó, V. García-Cañas, J.A. Ferragut, A. Cifuentes, Electrophoresis 33 (2012) 2328.

[24] R. Mandal, A.C. Guo, K.K. Chaudhary, P. Liu, F.S. Yallou, E. Dong, F. Aziat, D.S. Wishart, Genome Med. 4 (2012) 38.

[25] C. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, Anal. Chem. 78 (2006) 779.

[26] R. Tautenhahn, C. Böttcher, S. Neumann, BMC Bioinformatics 9 (2008) 504.

[27] R.A. Scheltema, A. Jankevics, R.C. Jansen, M.A. Swertz, R. Breitling, Anal. Chem. 83 (2011) 2786.

[28] D.J. Creek, A. Jankevics, K.E. Burgess, R. Breitling, M.P. Barrett, Bioinformatics 28 (2012) 1048.

[29] P.J. Martín-Álvarez, in: M.V. Moreno-Arribas, M.C. Polo (Eds.) Wine Chemistry and Biochemistry, Springer, New York, 2009, pp 677.

[30] A. Puerta, J.C. Díez-Masa, P.J. Martín-Álvarez, J.L. Martín-Ventura, C. Barbas, J. Tuñón, J. Egido, M. de Frutos, Analyst 136 (2011) 816.

[31] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, Nucleic Acids Res. 36 (2008) D480.

[32] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, Nucleic Acids Res. 37 (2009) D603.

[33] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, L.A. Mueller, Q. Ong, S. Paley, A. Pujar, A.G. Shearer, M. Travers, D. Weerasinghe, P. Zhang, P.D. Karp, Nucleic Acids Res. 36 (2007) D623.

[34] M. Sud, E. Fahy, D. Cotter, A. Brown, E.A. Dennis, C.K. Glass, A.H.Jr. Merrill, R.C. Murphy, C.R. Raetz, D.W. Russell, S. Subramaniam, Nucleic Acids Res. 35 (2006) D527.

[35] X. Liu, Q.H. Zhang, G.H. Yi, Mol. Cell Biochem. 363 (2012) 21.

[36] A.A. Farooqui, J. Alzheimers Dis. 30 (2012) S163.

[37] A. Dobolyi, G. Juhasz, Z. Kovacs, J. Kardos, Curr. Top. Med. Chem. 11 (2011) 1058.

[38] C.L. Pan, A. Gupta, H. Prentice, J.Y. Wu, J. Biomed. Sci. 17 (2010) S18.

[39] H.N. Noristani, A. Verkhratsky, J.J. Rodriguez, Aging Cell 11 (2012) 810.

[40] R. Cazzola, M. Rondanelli, M. Faliva, Exp. Gerontol. 47 (2012) 974.

[41] L. Szalardy, P. Klivenyi, D. Zadori, F. Fulop, J. Toldi, L. Vecsei, Curr. Med. Chem. 19 (2012) 1899.

[42] I. Degrell, F. Niklasson, Arch. Gerontol. Geriat. 7 (1988) 173.

[43] A. Algotsson, B. Winblad, Acta Neurol. Scand. 115 (2007) 403.

[44] C. Czech, P. Berndt, K. Busch, O. Schmitz, J. Wiemer, V. Most, H. Hampel, J. Kastler, H. Senn, PLOS ONE 7 (2012) e31501.

[45] Z. Kovacs, G. Juhasz, M. Palkovits, A. Dobolyi, K.A. Kekesi, Curr. Top. Med. Chem. 11 (2011) 1012.

[46] I. Santa-María, F. Hernández, M.A. Smith, G. Perry, J. Ávila, F.J. Moreno, Mol. Cell Biochem. 278 (2005) 203.

[47] K. Blennow, A. Wallin, C.G. Gottfries, A. Lekman, I. Karlsson, I. Skoog, L. Svennerholm, Neurobiol. Aging 13 (1992) 107.

[48] M. Sjogren, L. Minthon, U. Passant, K. Blennow, A. Wallin, Neurobiol. Aging 19 (1998) 379.

[49] N. Wszelaki, M.F. Melzig, Neurotoxicology 33 (2012) 424.

[50] W. Maruyama, T. Abe, H. Tohgi, P. Dostert, M. Naoi, Ann. Neurol. 40 (1996) 119.

[51] T.P. Roosild, S. Castronovo, A. Villoso, A. Ziemba, G. Pizzorno, J. Struct. Biol. 176 (2011) 229.

[52] M. Akbari, M. Otterlei, J. Pena-Diaz, H.E. Krokan, Neuroscience 145 (2007) 1201.

[53] G. Burnstock, U. Krugel, M.P. Abbracchio, P. Illes, Prog. Neurobiol. 95 (2011) 229.

[54] M. León-Otegui, R. Gómez-Villafuertes, J.I. Díaz-Hernández, M. Díaz-Hernández, M.T. Miras-Portugal, J. Gualix, FEBS Lett. 585 (2011) 2255.

[55] B. Plecko, C. Hikel, G.C. Korenke, B. Schmitt, M. Baumgartner, F. Baumeister, C. Jakobs, E. Struys, W. Erwa, S. Stöckler-Ipsiroglu, Neuropediatrics 36 (2005) 200.

[56] M.A. Ávila, E.R. García-Trevijano, S.C. Lu, F.J. Corrales, J.M. Mato, Int. J. Biochem. Cell Biol. 36 (2004) 2125.

APPENDIX A: Supplementary Data

A NEW METABOLOMIC WORKFLOW FOR EARLY DETECTION OF ALZHEIMER'S DISEASE

Clara Ibáñez[1], Carolina Simó[1,*], Dinesh K. Barupal[2], Oliver Fiehn[2], Miia Kivipelto[3], Angel Cedazo-Mínguez[3], Alejandro Cifuentes[1,*]

[1]Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain.

[2]University of California Davis, Genome Center, 451 E. Health Sci. Dr., Davis, CA 95616, USA.

[3]Karolinska Institute, NVS Department, KI-Alzheimer's Disease Research Center, 14186 Stockholm, Sweden.

Authors for the correspondance

*Address: Laboratory of Foodomics, CIAL (CSIC), Nicolas Cabrera 9, 28049 Madrid, Spain.

Carolina Simó:

e-mail: c.simo@csic.es. Telephone:+34-91-0017947. Fax: +34-91-0017905.

Alejandro Cifuentes:

e-mail: a.cifuentes@csic.es. Telephone: +34-91-0017955. Fax: +34-91-0017905.

Table of Contents

This supporting information file includes additional information as described in the text of the main article:

Experimental Section: Subjects, CSF sampling and metabolite extraction, Chemicals, Metabolomic analysis, Data processing, Multivariate statistical analysis, Metabolite identification and related pathways.

Tables: A-1, A-2, A-3, A-4, A-5, A-6, A-7

Figures: A-1.

2. MATERIALS AND METHODS

2.1 Subjects

The use of CSF samples for research was approved by the patients by written consent and by the ethics committee of the Karolinska University Hospital. This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethics committee of the Karolinska University Hospital, protocol number 2007/1:6 and date 2007-06-20. The methods used for clinical diagnosis of the patients included medical inspection, brain imaging, electroencephalography, blood and CSF screening tests (including total tau, phospho-tau, and Aβ 1-42 tests) and a detailed neuropsychological examination. MCI-S patients were not demented and had (self and/or an informant) reported cognitive decline and impairment on objective cognitive tasks. MCI-AD patients preserved basic activities of daily living and minimal impairment in complex instrumental functions. Dementia and AD were diagnosed according to DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) and NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association) criteria. Patients with psychiatric disorders (i.e. depression, alcohol abuse) or other conditions (i.e. diabetes, brain tumors, normal pressure hydrocephalus) at the time of CSF sampling were not included in this study.

2.2. CSF sampling and metabolite extraction

Lumbar puncture was performed in the sitting position in fasting subjects. CSF samples were obtained from L3/L4 or L4/L5 interspaces after local anesthetic infiltration in the skin. 10 mL were collected in polypropylene tubes after disposal of the first milliliter. No sample contained more than 500 erythrocytes per microliter of CSF. Samples were gently mixed to avoid gradient effects and centrifuged at 2000×g for 10 min to eliminate cells and insoluble material. Supernatants were immediately aliquot and stored at -80 °C for pending biochemical and metabolomic analyses.

2.3. Chemicals

Formic acid was from Riedel-de Haën (Seelze, Germany), acetic acid was from Scharlau (Barcelona, Spain), and ammonia was from Merck Millipore (Barcelona, Spain). Tyramine and DL-Methionine sulfone were selected as internal standards and

purchased from Sigma-Aldrich (St. Louis, MO, USA). Metabolite standards mixture for quality control assessment was composed of N-acetylputrescine, Cytidine, Adenosine, Methylthioadenosine, S-Adenosyl-homocysteine, Oxidized glutathione at concentrations of 18.5μM and L-Valine, L-Methionine, L-Tyrosine, Aminoadipic acid, L-Threonine, L-Tryptophan, DL-Hydroxyl-lysine, L-Anserine, Creatinine, L-Serine, L-Proline, L-leucine, Methyl-L-Histidine at concentrations of 0.375 mM, and purchased from Sigma-Aldrich. Acetonitrile and water were of MS grade and purchased from Labscan (Gliwice, Poland).

2.4. Metabolomic analysis

HILIC is a complementary analytical mode to the more common RP/LC, due to the ability of HILIC to separate more hydrophilic metabolites. Retention mechanisms in HILIC are a combination of hydrophilic interaction, ion-exchange, and reversed-phase retention that result in enhanced retention of polar analytes.

Internal mass calibration of the instrument was carried out using an AJS ESI source with an automated calibrant delivery system. The reference compound solution for internal mass calibration of the Q/TOF mass spectrometer containing 5 μM of purine ($[C_5H_5N_4]+$ at 121.050873 m/z) and 2.5 μM HP-0921, hexakis(1H,1H,3H-tetrafluoropropoxy) phosphazine ($[C_{18}H_{19}O_6N_3P_3F_{24}]+$ at 922.009798 m/z) in acetonitrile-water (95:5, v/v) was also from Agilent. External calibration of the TOF MS was carried out using a commercial mixture from Agilent with next m/z values: 118.086255, 322.048121, 622.028960, 922.009798, 1221.990637 and 1521.971475.

The quality of both chromatographic methods was assessed by the injection of a metabolite standards mixture containing molecules normally found in biological samples. The quality control mixture was prepared with standards presenting different nature, and distributed along the chromatogram (composition is shown in section 2.3. of this supporting information). Repeated injection of randomized samples (two replicates) together with the injection of standards mixture each fourth injection was performed to avoid the problem of large correlations occurring by chance when the number of samples is low (less than 100) and to assess the instrument measurement deviation [1]. In addition, two blank runs for RP method and three for HILIC method were injected before each sample.

2.5. Data processing

Peak detection method used in XCMS package5 was the CentWave algorithm, which finds a region of interest based on the expected mass accuracy and chromatographic peak width. The package mzMatch.R groups the features and performs noise filtering (intensity, peak shape, irreproducible peaks and ESI artifacts including isotopes, adducts and ion-source fragments), gap filling and related peaks annotation. IDEOM is a Microsoft Excel template with a collection of VBA macros that helps automated data processing of high resolution LC-MS data from untargeted metabolomics studies using a graphical user interface (GUI). In this work IDEOM was used to automatically check sample consistency in terms of standard deviation of the internal standards and to examine and revise the detected features. Filtering was then carried out to ensure ions with a high quality: (i) peaks below 3x intensity from the blanks were removed, (ii) peaks not found in at least a 75% of the samples of the same classification group or with a high variability within the same group (with a value of median/average > 1.5) were removed. The resulting output data table of high quality time-aligned detected peaks (also called "variables" along the text), with their corresponding migration time, molecular weight and peak area obtained for each sample, was used to perform a multivariate statistical analysis. Data processing was carried out using a standard HP computer Z400 (2.8GHz, 4 GB memory) running Windows 7. Additional details regarding XCMS, mzMatch and IDEOM are summarized in the supporting information. Data processing was carried out using a standard HP computer Z400 (2.8GHz, 4 GB memory) running Windows 7.

2.6. Multivariate statistical analysis

PPV can be defined as the ratio of true positives in relation to all the positive test results (true and false positives). The NPV is the ratio of true negatives in relation to all the negative test results (true and false negatives). Sensitivity can be defined as the ability of a diagnostic tool to detect the positive test results in relation to the total of true positives. A test with a high sensitivity has a low type II error rate or low number of false negative determinations. Specificity can be defined as the ability of a diagnostic tool to detect the negative test results in relation to the total of true negatives. A test with a high specificity has a low type I error rate or low number of false positive results.

2.7. Metabolite identification and related pathways

When isomers existed for a given formula, metabolite identification was sorted giving preference to metabolites from central metabolic pathways in KEGG, metabolites already found in CSF and number of databases containing each metabolite. When available, co-injection of standards with the sample was performed to confirm the tentatively identified metabolites.

SUPPORTING REFERENCES

[1] E. Zelena, W.B. Dunn, D. Broadhurst, S. Francis-McIntyre, K.M. Carroll, P. Begley, S. O'Hagan, J.D. Knowles, A. Halsall, HUSERMET Consortium I.D. Wilson, D.B. Kell, Anal. Chem. 81 (2009) 1357.

SUPPORTING TABLES

Table A-1. Summary of the parameters applied for data processing in this work.

XCMS	mzMatch	IDEOM
Method: centWave algorithm 25 ppm; peakwidth 3 – 65 (s); S/N 3; prefilter points 3; prefilter intensity 300; integrate 1; mzdiff 0.015 PeakML ionisation "detect", addscans 2.	Rtwindow 6; ppm 35; minintensity 300; mindetections 9. PeakML.GapFiller: ionisation "detect"; ppm 5; rtwindow 3.	Identification: ppm 10; Preferred database: Central KEGG. Search Adducts: double charged, Na^+, NH_3, HCOOH Internal standards: Tyramine and DL-Methionine sulfone.

Table A-2. LDA classification matrix of the samples. Observed and predicted classifications (with the probabilities of classification for each group) are shown in rows and columns respectively. One sample from MCI-AD group (sample 75) was misclassified as control group while sample 34 was correctly classified.

Classification group	% Correct Classification	**Control** (p = 0.28)	**MCI-S** (p = 0.28)	**MCI-AD** (p = 0.16)	**AD** (p = 0.28)
Control	100 %	20 + 1[a]	0	0	0
MCI-S	100 %	0	21	0	0
MCI-AD	91.7 %	1	0	11	0
AD	100 %	0	0	0	21
TOTAL	**98.7 %**	22	21	11	21

[a]Outlier from control group (sample 34).

Table A-3. Binary LDA classification models for all the group comparison. Sample groups compared in each case are marked in the first column. Calculated sensitivity, specificity, positive (PPV) and negative (NPV) predictive values are also calculated in each case.

Group comparisons	Number of significantly different (p<0.05) metabolites selected for the classification (in parenthesis ID of significant metabolites)	% Correct classification	% Sensitivity	% Specificity	% PPV	% NPV
Control vs. MCI-S	**4** (5, 2, 4, 18)	**95.2%** 95.2% (Control), 95.2% (MCI-S)	95.2	95.2	95.2	95.2
Contol vs. MCI-AD	**5** (6, 13, 4, 5, 19)	**100%** 100% (Control), 100% (MCI-AD)	100	100	100	100
Control vs. AD	**4** (6^{a}, 9, 20, 1)	**97.6%** 100% (Control), 95.2% (AD)	95.2	100	100	95.5
MCI-S vs. MCI-AD	**8** (21, 2, 4, 3, 22, 13, 23, 6)	**100%** 100% (MCI-S), 100% (MCI-AD)	100	100	100	100
MCI-S vs AD	**6** (1, 7, 10, 24, 3, 5)	**97.6%** 95.2% (MCI-AD), 100% (AD)	100	95.2	95.2	100
MCI-AD vs. AD	**3** (1, 19, 22)	**100%** 100% (MCI-AD), 100% (AD)	100	100	100	100

[a] More than one compound (adduct, complex…) corresponding to the same m/z value (the same metabolite).

Metabolite IDs order are given according to the forward stepwise LDA significance.

Table A-4. LOO-CV classification matrix of the samples: Observed and predicted classifications (with the probabilities of classification for each group) are shown in rows and columns respectively.

Classification group	% Correct classification	**Control** (p = 0.28)	**MCI-S** (p = 0.28)	**MCI-AD** (p = 0.16)	**AD** (p = 0.28)
Control	80.0 %	16	1	1	2
MCI-S	90.5 %	1	19	0	1
MCI-AD	75.0 %	0	1	9	2
AD	90.5 %	1	0	1	19
TOTAL	**85.1 %**	18	21	11	24

Table A-5. Mean values (±SE) of the 17 metabolites selected by the forward stepwise LDA, in the four groups of samples. The standard error (SE) was calculated according to the next formula: SE=SD/√n.; where SD is the standard deviation and n is the number of observations.

LDA significance	Tentative metabolites	Average peak area			
		Control	**MCI-S**	**MCI-AD**	**AD**
1	Uracil	9575 ± 1817.0	14213 ± 1576.1	18517 ± 1256.4	4930 ± 1096.7
2	Xanthine	53306 ± 2636.1	35748 ± 4377.6	71038 ± 6403.9	50553 ± 2968.9
3	Uridine	25267 ± 641.5	27333 ± 283.9	27597 ± 202.7	28078 ± 295.3
4	Tyrosil-serine/ Serinyl-tyrosine	27525 ± 310.8	27691 ± 356.9	25426 ± 646.9	28107 ± 337.0
5	Methylsalsolinol	27603 ± 421.5	30497 ± 368.7	29415 ± 451.7	28816 ± 412.2
6	Nonanoylglycine	18463 ± 428.1	19539 ± 484.3	22457 ± 772.8	18740 ± 380.3
7	Dopamine-quinone	3284 ± 329.4	2805 ± 261.1	3858 ± 542.6	4278 ± 321.3
8	Caproic acid	46941 ± 3595.5	54713 ± 2402.4	60695 ± 1350.3	51412 ± 3246.2
9	Vanylglycol	3543 ± 283.4	2719 ± 218.7	3365 ± 451.9	2099 ± 200.4
10	Histidine	1324 ± 42.3	1416 ± 39.4	1238 ± 19.4	1303 ± 36.9
11	Pipecolic acid	6137 ± 89.9	5748 ± 108.9	5736 ± 99.0	6255 ± 112.4
12	Hydroxyphosphinyl-pyruvate	53306 ± 2636.1	53462 ± 2795.7	71038 ± 6403.9	50268 ± 3064.9
13	Creatinine	15713 ± 1781.7	11248 ± 2058.5	7742 ± 2092.3	7443 ± 1612.6
14	Taurine	4216 ± 443.7	5229 ± 431.1	6028 ± 512.9	3859 ± 278.9
15	C16 Sphingosine-1-phosphate	2461 ± 404.1	3281 ± 626.5	4848 ± 758.9	2435 ± 282.4
16	Tryptophan	2120 ± 97.5	1929 ± 90.7	1723 ± 109.6	2209 ± 88.5
17	5'-Methylthioadenosine	3821 ± 126.6	4364 ± 106.7	4309 ± 102.1	4069 ± 158.3

Table A-6. Results from the multiple comparison procedures Scheffe, LSD and Games-Howell. Significant differences at p<0.05 are bolded.

Metabolite	Group	vs. Group	Significance level		
			Scheffe	**LSD**	**Games-Howell**
1	Control	MCI-S	0.095	**0.012**	0.144
		MCI-AD	**0.002**	**p<0.001**	**0.001**
		AD	0.282	0.052	0.245
	MCI-S	Control	0.095	**0.012**	0.144
		MCI-AD	0.341	0.069	0.165
		AD	**p<0.001**	**p<0.001**	**p<0.001**
	MCI-AD	Control	**0.002**	**p<0.001**	**0.001**
		MCI-S	0.341	0.069	0.165
		AD	**p<0.001**	**p<0.001**	**p<0.001**
	AD	Control	0.282	0.052	0.245
		MCI-S	**p<0.001**	**p<0.001**	**p<0.001**
		MCI-AD	**p<0.001**	**p<0.001**	**p<0.001**
2	Control	MCI-S	**0.029**	**0.003**	**0.014**
		MCI-AD	**0.03**	**0.003**	0.065
		AD	0.993	0.77	0.978
	MCI-S	Control	**0.029**	**0.003**	**0.014**
		MCI-AD	**p<0.001**	**p<0.001**	**0.001**
		AD	**0.049**	**0.005**	**0.039**
	MCI-AD	Control	**0.03**	**0.003**	0.065
		MCI-S	**p<0.001**	**p<0.001**	**0.001**
		AD	**0.013**	**0.001**	**0.047**
	AD	Control	0.993	0.77	0.978
		MCI-S	**0.049**	**0.005**	**0.039**
		MCI-AD	**0.013**	**0.001**	**0.047**

Table A-6 *(cont.)*

Metabolite	Group	vs. Group	Significance level		
			Scheffe	**LSD**	**Games-Howell**
3	Control	MCI-S	**0.005**	**p<0.001**	**0.032**
		MCI-AD	**0.008**	**0.001**	**0.012**
		AD	**p<0.001**	**p<0.001**	**0.003**
	MCI-S	Control	**0.005**	**p<0.001**	**0.032**
		MCI-AD	0.984	0.69	0.873
		AD	0.627	0.19	0.28
	MCI-AD	Control	**0.008**	**0.001**	**0.012**
		MCI-S	0.984	0.69	0.873
		AD	0.912	0.469	0.544
	AD	Control	**p<0.001**	**p<0.001**	**0.003**
		MCI-S	0.627	0.19	0.28
		MCI-AD	0.912	0.469	0.544
4	Control	MCI-S	0.988	0.72	0.979
		MCI-AD	**0.014**	**0.001**	**0.049**
		AD	0.727	0.256	0.574
	MCI-S	Control	0.988	0.72	0.979
		MCI-AD	**0.005**	**p<0.001**	**0.031**
		AD	0.885	0.423	0.832
	MCI-AD	Control	**0.014**	**0.001**	**0.049**
		MCI-S	**0.005**	**p<0.001**	**0.031**
		AD	**0.001**	**p<0.001**	**0.009**
	AD	Control	0.727	0.256	0.574
		MCI-S	0.885	0.423	0.832
		MCI-AD	**0.001**	**p<0.001**	**0.009**
5	Control	MCI-S	**p<0.001**	**p<0.001**	**p<0.001**
		MCI-AD	0.063	**0.007**	**0.035**
		AD	0.205	**0.034**	0.193
	MCI-S	Control	**p<0.001**	**p<0.001**	**p<0.001**
		MCI-AD	0.434	0.101	0.273
		AD	**0.034**	**0.003**	**0.021**
	MCI-AD	Control	0.063	**0.007**	**0.035**
		MCI-S	0.434	0.101	0.273
		AD	0.838	0.36	0.762
	AD	Control	0.205	**0.034**	0.193
		MCI-S	**0.034**	**0.003**	**0.021**
		MCI-AD	0.838	0.36	0.762

Table A-6 *(cont.)*

Metabolite	Group	vs. Group	Significance level		
			Scheffe	**LSD**	**Games-Howell**
6	Control	MCI-S	0.499	0.126	0.41
		MCI-AD	**p<0.001**	**p<0.001**	**0.002**
		AD	0.989	0.726	0.978
	MCI-S	Control	0.499	0.126	0.41
		MCI-AD	**0.004**	**p<0.001**	**0.022**
		AD	0.682	0.224	0.57
	MCI-AD	Control	**p<0.001**	**p<0.001**	**0.002**
		MCI-S	**0.004**	**p<0.001**	**0.022**
		AD	**p<0.001**	**p<0.001**	**0.003**
	AD	Control	0.989	0.726	0.978
		MCI-S	0.682	0.224	0.57
		MCI-AD	**p<0.001**	**p<0.001**	**0.003**
7	Control	MCI-S	0.744	0.269	0.625
		MCI-AD	0.815	0.335	0.841
		AD	0.259	**0.047**	0.199
	MCI-S	Control	0.744	0.269	0.625
		MCI-AD	0.286	0.053	0.332
		AD	**0.021**	**0.002**	**0.005**
	MCI-AD	Control	0.815	0.335	0.841
		MCI-S	0.286	0.053	0.332
		AD	0.893	0.436	0.909
	AD	Control	0.259	**0.047**	0.199
		MCI-S	**0.021**	**0.002**	**0.005**
		MCI-AD	0.893	0.436	0.909
8	Control	MCI-S	0.205	**0.034**	0.192
		MCI-AD	**0.027**	**0.003**	**0.004**
		AD	0.606	0.178	0.662
	MCI-S	Control	0.205	**0.034**	0.192
		MCI-AD	0.653	0.205	0.155
		AD	0.877	0.411	0.846
	MCI-AD	Control	**0.027**	**0.003**	**0.004**
		MCI-S	0.653	0.205	0.155
		AD	0.278	0.051	0.062
	AD	Control	0.606	0.178	0.662
		MCI-S	0.877	0.411	0.846
		MCI-AD	0.278	0.051	0.062

Table A-6 *(cont.)*

Metabolite	Group	vs. Group	Significance level		
			Scheffe	LSD	Games-Howell
9	Control	MCI-S	0.144	**0.021**	0.101
		MCI-AD	0.964	0.601	0.975
		AD	**0.002**	**p<0.001**	**0.001**
	MCI-S	Control	0.144	**0.021**	0.101
		MCI-AD	0.507	0.13	0.584
		AD	0.402	0.089	0.174
	MCI-AD	Control	0.964	0.601	0.975
		MCI-S	0.507	0.13	0.584
		AD	**0.036**	**0.004**	0.089
	AD	Control	**0.002**	**p<0.001**	**0.001**
		MCI-S	0.402	0.089	0.174
		MCI-AD	**0.036**	**0.004**	0.089
10	Control	MCI-S	0.514	0.133	0.523
		MCI-AD	0.465	0.112	0.182
		AD	0.94	0.529	0.935
	MCI-S	Control	0.514	0.133	0.523
		MCI-AD	**0.039**	**0.004**	**0.002**
		AD	0.191	**0.03**	0.173
	MCI-AD	Control	0.465	0.112	0.182
		MCI-S	**0.039**	**0.004**	**0.002**
		AD	0.759	0.282	0.417
	AD	Control	0.94	0.529	0.935
		MCI-S	0.191	**0.03**	0.173
		MCI-AD	0.759	0.282	0.417
11	Control	MCI-S	0.087	**0.011**	0.055
		MCI-AD	0.157	**0.023**	**0.037**
		AD	0.864	0.393	0.829
	MCI-S	Control	0.087	**0.011**	0.055
		MCI-AD	1,000	0.942	1,000
		AD	**0.008**	**0.001**	**0.012**
	MCI-AD	Control	0.157	**0.023**	**0.037**
		MCI-S	1,000	0.942	1,000
		AD	**0.028**	**0.003**	**0.008**
	AD	Control	0.864	0.393	0.829
		MCI-S	**0.008**	**0.001**	**0.012**
		MCI-AD	**0.028**	**0.003**	**0.008**

Table A-6 *(cont.)*

Metabolite	Group	vs. Group	Significance level		
			Scheffe	**LSD**	**Games-Howell**
12	Control	MCI-S	0.993	0.771	0.983
		MCI-AD	**0.01**	**0.001**	0.065
		AD	0.984	0.692	0.966
	MCI-S	Control	0.993	0.771	0.983
		MCI-AD	**0.016**	**0.001**	0.097
		AD	0.918	0.481	0.868
	MCI-AD	Control	**0.01**	**0.001**	0.065
		MCI-S	**0.016**	**0.001**	0.097
		AD	**0.003**	**p<0.001**	**0.044**
	AD	Control	0.984	0.692	0.966
		MCI-S	0.918	0.481	0.868
		MCI-AD	**0.003**	**p<0.001**	**0.044**
13	Control	MCI-S	0.289	0.054	0.269
		MCI-AD	0.051	**0.006**	**0.023**
		AD	**0.012**	**0.001**	**0.004**
	MCI-S	Control	0.289	0.054	0.269
		MCI-AD	0.7	0.236	0.635
		AD	0.514	0.133	0.474
	MCI-AD	Control	0.051	**0.006**	**0.023**
		MCI-S	0.7	0.236	0.635
		AD	1,000	0.919	0.999
	AD	Control	**0.012**	**0.001**	**0.004**
		MCI-S	0.514	0.133	0.474
		MCI-AD	1,000	0.919	0.999
14	Control	MCI-S	0.187	**0.03**	0.199
		MCI-AD	**0.024**	**0.002**	**0.027**
		AD	0.994	0.774	0.989
	MCI-S	Control	0.187	**0.03**	0.199
		MCI-AD	0.652	0.205	0.637
		AD	0.095	**0.012**	0.054
	MCI-AD	Control	**0.024**	**0.002**	**0.027**
		MCI-S	0.652	0.205	0.637
		AD	**0.011**	**0.001**	**0.008**
	AD	Control	0.994	0.774	0.989
		MCI-S	0.095	**0.012**	0.054
		MCI-AD	**0.011**	**0.001**	**0.008**

Table A-6 *(cont.)*

Metabolite	Group	vs. Group	Significance level		
			Scheffe	**LSD**	**Games-Howell**
15	Control	MCI-S	0.77	0.291	0.759
		MCI-AD	0.053	**0.006**	0.069
		AD	0.999	0.88	0.997
	MCI-S	Control	0.77	0.291	0.759
		MCI-AD	0.285	0.053	0.401
		AD	0.671	0.217	0.613
	MCI-AD	Control	0.053	**0.006**	0.069
		MCI-S	0.285	0.053	0.401
		AD	**0.034**	**0.003**	**0.043**
	AD	Control	0.999	0.88	0.997
		MCI-S	0.671	0.217	0.613
		MCI-AD	**0.034**	**0.003**	**0.043**
16	Control	MCI-S	0.385	0.083	0.328
		MCI-AD	**0.048**	**0.005**	**0.03**
		AD	0.982	0.68	0.976
	MCI-S	Control	0.385	0.083	0.328
		MCI-AD	0.586	0.167	0.483
		AD	0.185	**0.029**	0.137
	MCI-AD	Control	**0.048**	**0.005**	**0.03**
		MCI-S	0.586	0.167	0.483
		AD	**0.017**	**0.002**	**0.01**
	AD	Control	0.982	0.68	0.976
		MCI-S	0.185	**0.029**	0.137
		MCI-AD	**0.017**	**0.002**	**0.01**
17	Control	MCI-S	**0.023**	**0.002**	**0.008**
		MCI-AD	0.117	**0.016**	**0.019**
		AD	0.499	0.126	0.528
	MCI-S	Control	**0.023**	**0.002**	**0.008**
		MCI-AD	0.995	0.789	0.982
		AD	0.426	0.097	0.421
	MCI-AD	Control	0.117	**0.016**	**0.019**
		MCI-S	0.995	0.789	0.982
		AD	0.717	0.248	0.586
	AD	Control	0.499	0.126	0.528
		MCI-S	0.426	0.097	0.421
		MCI-AD	0.717	0.248	0.586

Table A-7. Tentative identification and related pathways of the metabolites selected in the linear discriminant analysis.

Metabolite ID	Tentative Identification	Monoisotopic Mass	Error ppm	HMDB ID	KEGG ID	Molecular Formula	KEGG Map	KEGG Pathway
1	Uracil[a]	112.027	1.26	HMDB00300	C00106	C4H4N2O2	hsa00240 hsa00410 hsa00770	Pyrimidine metabolism beta-Alanine metabolism Pantothenate and CoA biosynthesis
2	Xanthine	152.033	-0.45	HMDB00292	C00385	C5H4N4O2	hsa00230	Purine metabolism
3	Uridine	244.070	2.73	HMDB00296	C00299	C9H12N2O6	hsa00240	Pyrimidine metabolism
4	Tyrosyl-Serine / Serinyl-Tyrosine	209.069	-8.09	HMDB29114 / HMDB29051	---	C12H16N2O5	---	---
5	Methylsalsolinol	193.110	5.02	HMDB03626 / HMDB03892	---	C11H15NO2	---	---
6	Nonanoylglycine	215.152	-2.19	HMDB13279	C02055[b]	C11H21NO3	hsa00360	Phenylalanine metabolism
7	Dopamine quinone	151.063	5.11	HMDB12219	C17755	C8H9NO2	hsa00965	Betalain Biosynthesis
8	Caproic acid	116.084	-3.14	HMDB00535	C01585	C6H12O2	---	---
9	Vanylglycol	184.074	-0.12	HMDB01490	C05594	C9H12O4	hsa00350	Tyrosine metabolism
10	Histidine[a]	155.069	7.88	HMDB00177	C00135	C6H9N3O2	hsa00340 hsa00410 hsa00970	Histidine metabolism Beta-alanine metabolism Aminoacyl-tRNA biosynthesis
11	Pipecolic acid	129.079	-0.02	HMDB00716	C00408	C6H11NO2	hsa00310	Lysine degradation
12	Hydroxyphosphinyl-piruvate	151.988	9.25	---	C06368	C3H5O5P	hsa00330	Arginine and proline metabolism
13	Creatinine[a]	113.059	0.12	HMDB00562	C00791	C4H7N3O	hsa00330	Arginine and proline metabolism
14	Taurine[a]	125.015	0.29	HMDB00251	C00245	C2H7NO3S	hsa00120 hsa00430 hsa02010 hsa0408	Primary bile acid biosynthesis Taurine and hypotaurine metabolism ABC Transporters: Mineral and organic transporter Neuroactive ligand-receptor interaction: Glutamate receptor (GLR)
15	Sphingosine-1-phosphate	351.218	8.01	LMSP01050005[c]	---	C16H34NO5P	---	---

Table A-7 (cont.)

Metabolite ID	Tentative Identification	Monoisotopic Mass	Error ppm	HMDB ID	KEGG ID	Molecular Formula	KEGG Map	KEGG Pathway
16	Tryptophan[a]	204.090	-0.72	HMDB00929	C00078	C11H12N2O2	hsa00260 hsa00380 hsa00400 hsa00970 hsa04726 hsa04974 hsa04978	Glycine, serine and threonine metabolism Tryptophan metabolism Phenylalanine, tyrosine and tryptophan biosynthesis Aminoacyl-tRNA biosynthesis Serotonergic synapse Protein digestion and absorption Mineral absorption: Sodium absorption to interact with "solute carrier family 6 (neurotransmitter transporter) member 19" = "SLC6A19"
17	Methylthioadenosine[a]	297.090	2.12	HMDB01173	C00170	C11H15N5O3S	hsa00270	Cysteine and methionine metabolism
18	Methylhistidine[a]	169.085	2.25	HMDB00001 / HMDB00479	C01152	C7H11N3O2	hsa00340	Histidine metabolism
19	Histidinyl-Lysine / Lysyl-Histidine	283.164	2.36	HMDB28890 / HMDB28953	---	C12H21N5O3	---	---
20	Octylamine	129.152	0.02	---	C01740	C8H19N	---	---
21	6-deoxygalactose	164.069	1.23	HMDB12327	C01019 / C02781	C6H12O5	hsa00051 hsa00520	Fructose and mannose metabolism Amino sugar and nucleotide sugar metabolism
22	PG(15:1(9Z)/0:0)	468.249	2.41	LMGP04050018[c]	---	C21H41O9P	---	---
23	Aspartyl-aspartate	248.064	5.18	HMDB28749	---	C8H12N2O7	---	---
24	Homovanillic acid	182.173	4.27	HMDB00118	---	C9H10O4	hsa00350 hsa04728	Tyrosine metabolism Dopaminergic synapse

[a] Identification confirmed by commercial standard injection; [b] Acylglycine identifier; [c] Not HMDB ID available. LipidMaps ID show

SUPPORTING FIGURES
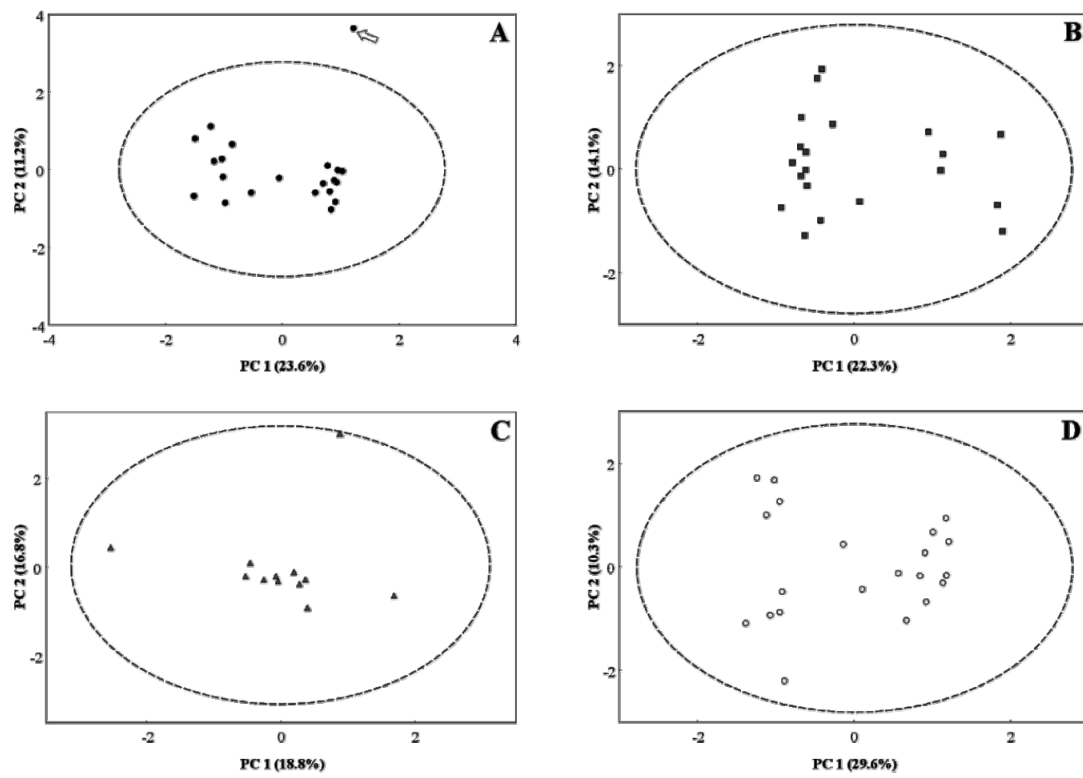
Figure A-1. Plot of the samples in the plane defined by the two first principal components (PC) representing the greatest variance in the data of control group (A), MCI-S group (B), MCI-AD group (C) and AD group (D). The percentages of the total variance explained by each component (PC1 and PC2) are also shown. 95% confidence ellipses are included in order to find possible outliers which are indicated by arrows.

4.6. CONCLUSIONES

En el **Capítulo 4** se ha realizado una exhaustiva revisión bibliográfica de todos los trabajos publicados hasta diciembre de 2012 relacionados con el estudio de la enfermedad de Alzheimer (AD), empleando diferentes estrategias metabolómicas no dirigidas. La Metabolómica ha demostrado una gran capacidad para la clasificación de muestras (tejidos, fluidos biológicos) de pacientes control (sin AD) y pacientes con AD. En general existe una conclusión general relacionada con la necesidad de aumentar el tamaño muestral en los estudios clínicos enfocados hacia la búsqueda de biomarcadores para el diagnóstico de la enfermedad. También es destacable la necesidad de ampliar los estudios de diagnóstico precoz de la enfermedad a pacientes con deterioro cognitivo leve con el fin de predecir cuáles de estos pacientes desarrollarán Alzheimer o bien permanecerán estables sin desarrollar la enfermedad. En este sentido se puede destacar que los trabajos de investigación presentados en esta Tesis (**Capítulo 5**, **Capítulo 6** y **Capítulo 7**) abordan por primera vez el estudio de muestras de CSF en estos grupos de pacientes con MCI.

En el **Capítulo 5** se ha desarrollado un método basado en MEKC-LIF quiral para la separación enantioselectiva de un grupo de aminoácidos: Gly, D/L-Arg, D/L-Ser, D/L-Leu, D/L-Ala, D/L-Gln, D/L-Glu, GABA, D/L-Lys y D/L-Asp. Este método permitió el análisis enantioselectivo, empleando β-ciclodextrinas (β-CD) como selector quiral, con una elevada sensibilidad (límite de detección entre 0.8 nM y 16.5 nM), eficiencia (hasta 703000 platos/m) y resolución en la separación de estos aminoácidos (de 2.6 a 9.5).

Este método se ha aplicado al análisis de muestras de CSF procedentes de pacientes con un diagnóstico inicial de deterioro cognitivo subjetivo (SCI), demencia incipiente (MCI) y AD. Durante los dos años inmediatamente posteriores a la recogida de las muestras se llevó a cabo un seguimiento clínico de los pacientes. Algunos de los pacientes, cuyo diagnóstico inicial era MCI, progresaron a AD (MCI-AD) mientras que otros presentaron MCI estable (MCI-S). Se observó una cantidad significativamente menor de L-Arg L-Lys, L-Glu y L-Asp en CSF procedentes de pacientes que presentaban o desarrollaron AD (AD y MCI-AD) con respecto a los que no

213

desarrollaron la enfermedad (SCI, MCI-S) en el trascurso del estudio. Por el contrario, se observó un aumento acusado en los niveles de GABA en sujetos que desarrollaron en algún momento AD (MCI-AD y AD), con la máxima concentración detectada en aquellos individuos diagnosticados inicialmente con la enfermedad (AD).

Este trabajo revela el potencial del método MEKC-LIF desarrollado para la detección de diferencias en algunos aminoácidos quirales en muestras de CSF procedentes de pacientes con diferentes grados de deterioro cognitivo y AD. Sin embargo es necesario extender el estudio a un número elevado de muestras para validar los resultados obtenidos en este trabajo.

En el **Capítulo 6** se ha empleado la técnica CE-TOF MS para el estudio metabolómico de muestras de CSF con diferentes grados de deterioro cognitivo y AD. Concretamente, se estudiaron los cuatro grupos de muestras que se indican a continuación: SCI (n=19), MCI-S (n=22), MCI-AD (n=9), AD (n=23). Además se analizó un grupo de muestras de diagnóstico desconocido (n=12) para la validación del método de clasificación.

Se detectaron un total de 160 metabolitos en las muestras de CSF. Entre los 160 metabolitos detectados, 10 destacaron como los más diferenciadores entre los cuatro estados cognitivos (SCI, MCI-S, MCI-AD y AD) proporcionando un 90.1% de clasificación correcta de las muestras en su grupo de diagnóstico.

Se observó además que aquellos sujetos que no desarrollaron AD en ningún momento del estudio (SCI y MCI-S) presentaban perfiles metabólicos muy similares, por lo que se diseñó otro modelo predictivo fusionando dichos grupos en uno sólo (non-AD). Como resultado del análisis lineal discriminante entre los tres grupos de pacientes (non-AD, MCI-AD y AD) el porcentaje de asignación correcta de los individuos en su grupo de diagnóstico aumentó considerablemente (97.2%) empleando los niveles de 14 metabolitos.

En este trabajo se realizaron todas las comparativas posibles entre grupos de diagnóstico distintintos observándose diferencias en la expresión de 71 metabolitos, de los que 61 se identificaron tentativamente y 9 fueron corroborados mediante la co-

inyección de la muestra con estándares comerciales (colina, arginina, histidina, dimetil-arginina, carnitina, creatina, valina, serina y prolina).

La colina, la arginina, la carnitina y la valina (así como dos compuestos no identificados) destacaron por estar involucrados tanto en la diferenciación entre los cuatro grupos de muestras originales (SCI, MCI-S, MCI-AD y AD) como en la comparativa posterior entre los tres grupos (non-AD, MCI-AD y AD).

Para corroborar la utilidad del modelo de clasificación creado entre los tres grupos de muestras (non-AD, MCI-AD y AD), los resultados estadísticos se sometieron a procesos de validación externa (clasificación del grupo de muestras de validación) e interna (validación cruzada, LOO-CV) obteniendo en ambos casos porcentajes por encima del 85% de clasificación correcta, lo que confirma el potencial del método de clasificación diseñado y de los potenciales biomarcadores propuestos en este estudio.

En relación con la necesidad del desarrollo de nuevas estrategias analíticas para el diagnóstico precoz de AD, en el **Capítulo 7** se muestra la utilidad de dos métodos analíticos, HILIC/UPLC-TOF MS y RP/UPLC-TOF MS para el análisis metabolómico de muestras de CSF.

Se detectaron 286 metabolitos (mediante HILIC/UPLC-TOF MS) y 238 metabolitos (mediante RP/UPLC-TOF MS) en 75 muestras de CSF procedentes de cuatro estadíos cognitivos diferentes (SCI, MCI-S, MCI-AD y AD). Como resultado de un análisis lineal discriminante (LDA), se seleccionaron diecisiete metabolitos por tener una mayor capacidad discriminante entre los grupos de diagnóstico. Con los niveles de estos 17 metabolitos se alcanzó un 98.7% de asignación correcta de las muestras en su grupo de clasificación y valores de sensibilidad y especificidad del método diagnóstico preliminar desarrollado superiores al 95%. Se obtuvieron valores de validación (LOO-CV) superiores al 85%, lo que corrobora la utilidad del método de clasificación desarrollado.

Tras la comparación entre todos los grupos de diagnóstico distintintos se detectaron diferencias en la expresión de 24 metabolitos, que se identificaron tentativamente y se asignaron a las rutas metabólicas en las que participan. Se pudo

corroborar la identidad tentativa del uracilo, la histidina, la creatinina, la taurina, el triptófano, la metiltioadenosina y la metilhistidina mediante la coinyección de la muestra con los estándares comerciales.

La histidina fue el único metabolito común tanto en la creación del modelo predictivo tras el análisis mediante CE-MS como el desarrollado tras el análisis mediante UPLC-MS mostrando de nuevo la complementariedad de ambas técnicas analíticas y la gran cobertura metabólica alcanzada mediante el uso de multiplataformas analíticas.

Los trabajos publicados en los tres últimos capítulos de esta Tesis se enmarcan en una etapa de descubrimiento ("*discovery phase*") dentro de los estudios clínicos cuya función es la de generar nuevas hipótesis que requieren una corroboración en una etapa posterior de validación ("*validation phase*") con un número muy superior de muestras (100-1000 muestras).

5. CONCLUSIONES GENERALES Y TRABAJO FUTURO

Las principales conclusiones del trabajo presentado en esta Tesis doctoral se enumeran a continuación:

1. Se han desarrollado y optimizado métodos analíticos útiles, reproducibles y eficientes basados en la hibridación de técnicas electroforéticas/cromatográficas y espectrometría de masas para el análisis metabolómico de muestras biológicas, en concreto, cultivos celulares y líquido cefalorraquídeo. Además, en función de los objetivos establecidos para cada estudio, se han elegido y diseñado las pautas más adecuadas para el procesamiento de los datos y su posterior análisis estadístico mediante herramientas bioinformáticas avanzadas.

2. Los métodos analíticos desarrollados se han aplicado con éxito al estudio de las diferencias metabolómicas debidas al tratamiento con un extracto de romero rico en polifenoles en modelos celulares de cáncer de colon humano y al estudio de distintos estados cognitivos relacionados con la progresión de la enfermedad de Alzheimer. Las diferencias observadas entre las distintas muestras biológicas se han relacionado con las rutas metabólicas afectadas, profundizando de este modo en los mecanismos moleculares afectados en cada caso.

3. La utilidad de la integración de los resultados procedentes de multiplataformas analíticas en estudios metabolómicos ha sido demostrada a lo largo de los capítulos de esta Tesis. Asimismo, los resultados metabolómicos se han integrado con los resultados de estudios de Transcriptómica y Proteómica siguiendo una aproximación Foodómica. A pesar del elevado número de trabajos publicados sobre polifenoles y su efecto en la salud, en nuestro conocimiento, en esta Tesis doctoral se presenta por vez primera el estudio de dicho efecto a nivel molecular integrando los resultados tras un análisis transcriptómico, proteómico y metabolómico.

4. El estudio, demostración y explicación a nivel molecular del efecto antiproliferativo del extracto de romero rico en polifenoles (carnosol y ácido carnósico, fundamentalmente) sobre células humanas modelo de cáncer de colon HT-29 se engloba dentro de un proyecto más amplio. En la actualidad se están llevando a cabo los primeros estudios ómicos del efecto de carnosol y ácido

carnósico puros sobre células HT-29, con el fin de corroborar los resultados obtenidos con el extracto rico en polifenoles, observar posibles sinergias, y en su caso, ampliar la información a nivel molecular obtenida mediante los estudios transcriptómicos, proteómicos y metabolómicos, vinculando así la actividad antiproliferativa a unos determinados compuestos.

5. En relación a los estudios sobre la enfermedad de Alzheimer, hasta la fecha la fase de demencia incipiente o MCI es la fase más temprana conocida de esta enfermedad. Teniendo en cuenta las evidencias científicas de la existencia de una fase presintomática de la AD, es necesario continuar y profundizar en las investigaciones de búsqueda de biomarcadores tempranos, para lo cual se prevé en un futuro el empleo de estrategias metabolómicas en estudios longitudinales orientados a encontrar marcadores biológicos que permitan predecir la aparición de la enfermedad antes de que se manifiesten los síntomas.

6. Los resultados metabolómicos obtenidos en esta Tesis doctoral sobre la detección de nuevos biomarcadores de AD relacionados con la evolución de la misma y la posibilidad de su detección en un estadio temprano, pueden permitir profundizar en el estudio de otras líneas de actuación (p. ej., el efecto de la dieta) que permitan detener o ralentizar su progresión. Recientes estudios sugieren que un tipo de nutrición basada en el patrón tradicional de dieta mediterránea podría prevenir o retrasar el deterioro cognitivo. Además cada vez hay más evidencias que proponen que los factores de riesgo cardiovascular contribuyen a los problemas cognitivos asociados a la edad así como a otros tipos de demencia como la AD y la demencia vascular. Por esto, una adecuada intervención nutricional a nivel clínico podría ayudar al tratamiento de esta enfermedad. Sin embargo, aunque las evidencias al respecto son prometedoras, es necesario profundizar en los mecanismos moleculares que intervienen en estos efectos potencialmente protectores, para lo cual, las tecnologías ómicas desempeñarán un papel fundamental.

GENERAL CONCLUSIONS AND FUTURE WORK

The following main conclusions are drawn from this PhD Thesis work:

1. Useful, reproducible and efficient analytical methodologies have been developed and optimized for metabolomic analysis of biological samples (namely, cell cultures and cerebrospinal fluid), based on the hybridization of electrophoretic/chromatographic techniques and high resolution mass spectrometry. In addition, appropriate bioinformatic protocols have been selected and designed according to the specificity of each study.

2. Metabolomic differences on colon cancer cells (HT-29 cell line) treated with a rich-polyphenols extract from rosemary, and those differences due to different cognitive status related to Alzheimer's disease, have been successfully investigated by the application of the developed analytical methodologies. Differences observed between samples have been correlated with altered metabolic pathways, and as a result, we have been able to deepen the current knowledge on the main molecular mechanisms implicated.

3. Integration of data obtained by the analytical metabolomic multiplatforms has been shown to be very useful in this PhD Thesis. Metabolomic data integration with results from other expression levels (trascriptome and proteome) has also been further tackled following a foodomics approach. Despite the effect of polyphenols on health have been already published in hundreds of scientific papers, the investigation on the effects of polyphenols at molecular level by the integration of Transcriptomics, Proteomics and Metabolomics has been addressed for the first time in this PhD Thesis work.

4. The investigation on the antiproliferative effects of a rich-polyphenol (mainly carnosic acid and carnosol) extract from rosemary against human colon cancer cells, is part of a more comprehensive project. At present, preliminary omic studies are being carried out to confirm the previously observed effects with the selected rosemary extract, using pure carnosol and carnosic acid and their potential synergic effects on HT-29 colon cancer cells. This study will increase our understanding on the molecular

mechanisms altered at transcriptomic, proteomic and metabolomic level. The goal will be to unequivocally assign the antiproliferative activity to given compounds.

5. Although recent scientific evidences have stated the existence of unknown pre-symptomatic Alzheimer's disease (AD) phases, up to now mild cognitive impairment (MCI) is the earliest known stage in AD progression. These facts together with the scarce knowledge about AD pathophysiology make imperative the research in new early AD biomarkers as done in this PhD Thesis. Biomarkers discovery in pre-symptomatic AD stages by metabolomic strategies, in large longitudinal studies, is foreseen as an essential approach in future AD research.

6. Metabolomic results for early AD biomarkers discovery as the ones obtained in this PhD Thesis can make easier to explore other research lines (i.e. diet effect) to stop or slow AD progression. Nutrition based on Mediterranean diet patterns has been recently suggested as a preventive or retarder factor for cognitive decline process. On the other hand, cardiovascular and cognitive disorders (age-related, AD-related, vascular dementia, etc.) are being increasingly associated each other. However, molecular mechanisms related to the preventive effects of diet ingredients are still not clear and for that, omic technologies will offer crucial results.

APÉNDICE I / APPENDIX I

PRODUCCIÓN CIENTÍFICA.
PAPERS GENERATED IN THIS PhD THESIS.

SCI papers

- Carolina Simó, **Clara Ibáñez**, Angeles Gómez-Martínez, Jose Antonio Ferragut, Alejandro Cifuentes*

  "*Is Metabolomics reachable? Different purification strategies of human colon cancer cells provide different CE-MS metabolite profiles*"

  Electrophoresis 2011, 32, 1765-1777.

- Shorena Samakashvili, **Clara Ibáñez**, Carolina Simó, Francisco Javier Gil-Bea, Bengt Winblad, Angel Cedazo-Mínguez, Alejandro Cifuentes*

  "*Analysis of chiral amino acids in cerebrospinal fluid samples linked to different stages of Alzheimer disease*"

  Electrophoresis 2011, 32, 2757-2764.

- **Clara Ibáñez**, Alberto Valdés, Virginia García-Cañas, Carolina Simó, Mustafa Celebier, Lourdes Rocamora-Reverte, Ángeles Gómez- Martínez, Miguel Herrero, María Castro, Antonio Segura-Carretero, Elena Ibáñez, José A. Ferragut, Alejandro Cifuentes*

  "*Global foodomics strategy to investigate the health benefits of dietary constituents*"

  Journal of Chromatography A 2012, 1248, 139-153.

- Alberto Valdés, Carolina Simó, **Clara Ibáñez**, Lourdes Rocamora, José Antonio Ferragut, Virginia García-Cañas*, Alejandro Cifuentes

  "*Effect of dietary polyphenols on K562 leukemia cells: A foodomics approach*"

  Electrophoresis 2012, 33, 2314-2327.

- **Clara Ibáñez**, Carolina Simó*, Virginia García-Cañas, Ángeles Gómez-Martínez, José A. Ferragut, Alejandro Cifuentes

  "*CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation*"

  Electrophoresis 2012, 33, 2328-2336.

- **Clara Ibáñez**, Carolina Simó, Pedro Martín-Alvarez, Miia Kivipelto, Bengt Winblad, Ángel Cedazo-Minguez, Alejandro Cifuentes*

  "*Towards a predictive model of Alzheimer's disease progression using capillary electrophoresis-mass spectrometry Metabolomics*".

  Analytical Chemistry 2012, 84, 8532-8540.

- **Clara Ibáñez**, Carolina Simó*, Alejandro Cifuentes

  "*Metabolomics in Alzheimer's disease research*".

  Electrophoresis, 2013 (in press)

- Alberto Valdés, **Clara Ibáñez**, Carolina Simó, Virginia García-Cañas*

  "*Recent transcriptomics advances and emerging applications in food science*".

  Trends in Analytical Chemistry, 2013 (in press).

- **Clara Ibáñez**, Carolina Simó*, Dinesh K. Barupal, Oliver Fiehn, Miia Kivipelto, Ángel Cedazo-Minguez, Alejandro Cifuentes*

  "*A new metabolomic workflow for early detection of Alzheimer's disease*".

  Journal of Chromatography A, 2013 (sent).

- **Clara Ibáñez**, Carolina Simó, Virginia García-Cañas, Alejandro Cifuentes*, María Castro-Puyana

  "*Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics*".

  Analitica Chimica Acta, 2013 (sent).

- Alberto Valdés, Carolina Simó, **Clara Ibáñez**, Virginia Garcia Cañas*

  "*Foodomics strategies for the analysis of transgenic foods*".

  Trends in Analytical Chemistry, 2013 (sent).

- **Clara Ibáñez**, Virginia García-Cañas, Alberto Valdés, Carolina Simó*

  "*Novel MS-based approaches and applications in food Metabolomics*".

  Trends in Analytical Chemistry, 2013 (sent).

Book chapters and No-SCI papers

- Mustafa Celebier, **Clara Ibáñez**, Carolina Simó, Alejandro Cifuentes
  "*A foodomics approach: CE-MS for comparative metabolomics of colon cancer cells treated with dietary polyphenols*". Chapter 15, In: Protein electrophoresis: methods and protocols, Edited by Biji Kurien and Hal Scofield. Series: Methods in Molecular Biology. Editorial: Springer, New York 2012, pp. 185-195.

- **Clara Ibáñez** and Carolina Simó
  "*MS-based metabolomics in nutrition and health research*" Chapter 9, In: Foodomics: Advanced mass spectrometry in modern food science and nutrition, Edited by Alejandro Cifuentes. Editorial: John Wiley & Sons, Inc., Hoboken, New Jersey 2013, pp. 245-270.

- Alberto Valdés, Virginia García-Cañas, Carolina Simó, Alejandro Cifuentes*, **Clara Ibáñez**
  "*Dietary components and their effect in colon cáncer: Foodomic evaluation using Transcriptomics and Metabolomics*".
  SECyTA bulletin, 2012, 33, 51-68.