



LC-MS BASED TARGETED AND UNTARGETED METABOLOMIC APPROACHES FOR THE IDENTIFICATION OF POTENTIAL BIOMARKERS IN PLASMA FROM PAEDIATRICS WITH CHRONIC KIDNEY DISEASE



SANDRA BENITO CID
2018

Esta Tesis ha sido realizada en el Departamento de Química Analítica de la Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU) y en el Servicio Central de Análisis de Alava (SGIker).

Parte del trabajo ha sido desarrollado en el grupo de Quimiometría del Departamento de Química Analítica de la Universidad de Radboud (Nijmegen, Países Bajos) durante una estancia de investigación de tres meses de duración (31 enero-1 mayo 2015), con el propósito de obtener el título de Doctorado Internacional por la UPV/EHU.

Previamente a la defensa de esta Tesis Doctoral, este trabajo ha sido evaluado por dos expertos pertenecientes a Centros de Investigación de la Unión Europea:

- ✓ **Dra. Raquel Gutiérrez Climente**, Université de Technologie de Compiègne (Compiègne, Francia).

- ✓ **Dra. Nuria de Diego Sánchez**, Palacký University (Olomouc, República Checa).

Agradecimientos / Acknowledgements

Al concluir un trabajo como es la tesis, a veces agradecida, a veces ardua, es cuando uno llega a ver el trabajo completo en perspectiva, y se da cuenta de que la magnitud de éste hubiera sido imposible sin el apoyo de numerosas personas e instituciones que me han brindado su apoyo.

En primer lugar, me gustaría agradecer a mis directoras, las doctoras Alicia Sánchez y Nora Unceta, por todo lo que me han enseñado y ayudado durante esta tesis. Nora, aún me recuerdo tocando a tu puerta con la intención de conocer este pequeño mundo que es la investigación a través de la beca de colaboración. Ha llovido bastante desde entonces, aunque el tiempo se me ha pasado volando... Agradecerte que me descubrieras la investigación y sobre todo tu paciencia y disposición. Ali, cuántos días a pie de equipo tirándonos de los pelos, revisando exceles que parecían no tener fin, saltando de alegría cuando un giro inesperado cambiaba nuestra suerte. Gracias por compartir tus conocimientos en metabolómica, LC-MS y mantenimiento de los equipos, y por tu paciencia con mis mails kilométricos.

Agradecer también a los profesores Ramón J. Barrio y Arantxa Goicolea por darme la oportunidad de formar parte de su grupo de Investigación, Metabolomips, poniendo a mi disposición los medios y el apoyo necesarios para ejecutar este trabajo.

Gracias a la Universidad del País Vasco (UPV/EHU), al Servicio Central de Análisis de Álava (SCAA) y al departamento de Química Analítica por las instalaciones, equipamiento, el soporte técnico y los medios ofrecidos, que han permitido la consecución de este trabajo.

Además, me gustaría agradecer la financiación recibida en forma de contrato predoctoral por parte del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (PRE_2013_1_899).

I gratefully acknowledge Prof. Lutgarde M.C. Buydens and her research group in Radboud University (Nijmegen, The Netherlands), for accepting me in their group and having a fruitful stay. Agradecer además, la financiación recibida por parte del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco en forma de beca de movilidad (EP_2016_1_0003) para realizar dicha estancia.

Por otro lado, dar las gracias al grupo de Metabolismo del Instituto de Investigación Sanitaria Biocruces por facilitarnos las muestras de pacientes que han permitido llevar a

Agradecimientos / Acknowledgements

cabo los estudios contenidos en este manuscrito, y en particular al Dr. Fernando Andrade por su disposición.

A mis compañeros de laboratorio, tanto a los actuales como a los que han tomado otros rumbos (algunos de ellos, además de compañeros, buenos amigos). No me quiero olvidar de nadie. Por vuestras aportaciones en lo profesional, pero también por el apoyo personal en los buenos y en los malos momentos. ¡Gracias por tanto!

A mis amigos fuera del círculo científico (mis "domingueros", gente de la pista de atletismo, mis niños de combat...), por ayudarme a evadirme para no perder el entusiasmo y tampoco el norte.

A ti Alberto, por estar ahí a lo largo de todos estos años, comprender lo que la tesis significaba para mí y apoyarme incondicionalmente en esta aventura. Gracias por mantenerte al pie del cañón para que mi día pudiera tener 48 h y ayudarme a convertir este sueño en realidad.

Por último, pero no menos importante, a mi familia y en especial a mis padres, Nieves y Justo. Por enseñarnos el valor del esfuerzo y apoyarnos siempre, desde la humildad y el sacrificio, para que tuviéramos el mejor futuro posible. Y a ti Nerea, por ser mucho más que una hermana. Por tu confianza, cariño y amistad. Non puoi immaginare quanto mi manchi. Os quiero.

ESKERRIK ASKO!

*"Un scientifique dans son laboratoire est non seulement un technicien:
il est aussi un enfant placé devant des phénomènes naturels qui l'impressionnent
comme des contes de fées"*

Marie Curie

*"We all have dreams. But in order to make dreams come into reality, it takes an
awful lot of determination, dedication, self-discipline, and effort".*

Jesse Owens



CONTENTS

Acronyms & Abbreviations	V
Resumen	XIII
Summary	XXI

Chapter I: Introduction

1.1 Chronic kidney disease. Definition	3
1.2 Etiopathology of CKD	3
1.3 CKD in adults and in paediatrics.....	6
1.4 CKD degree	7
1.5 Epidemiology of CKD	8
1.6 Risk factors for CKD progression and complications of CKD	10
1.7 CKD treatment	11
1.8 Diagnosis and evaluation of CKD.....	14
1.9 Biomarkers. Compounds of interest for the early detection and intervention on the disease	16
1.10 Bibliography.....	18

Chapter II: Objectives

2.1 Objectives	27
----------------------	----

Chapter III: Methodology

3.1 Introduction.....	31
3.2 Metabolomics	31
3.3 Analytical instrumentation.....	33
3.3.1 Nuclear magnetic resonance.....	33
3.3.2 Mass spectrometry.....	34
3.4 Statistical and chemometrical approach	47
3.4.1 Hierarchical clustering	50
3.4.2 Principal component analysis.....	50

CONTENTS

3.4.3	Classification methods	52
3.4.4	Sparse partial least squares discriminant analysis	56
3.5	Bibliography	57

Chapter IV: Targeted metabolomics analysis of amino acids and related compounds using LC-QTOF-MS

4.1	Introduction	65
4.2	Metabolic pathways suspicious of being affected in paediatrics with CKD: literature review	66
4.3	Targeted metabolomics workflow	69
4.4	Study's goal	70
4.5	Description of targeted metabolites	71
4.6	Materials and equipment	78
4.6.1	Standards and reagents	78
4.6.2	Instrumental analysis	79
4.6.3	Preparation of standard solutions	80
4.6.4	Subjects and sampling	81
4.7	Optimization of the analytical method	81
4.7.1	Selection of the chromatographic variables	84
4.7.2	Selection of the mass spectrometry variables	88
4.7.3	Optimization of sample treatment conditions	98
4.8	Analytical method evaluation	106
4.8.1	Calibration of the method and matrix effect	106
4.8.2	Limits of quantification, accuracy and precision	111
4.8.3	Recovery from the extraction process	113
4.8.4	Stability	113
4.9	Sample analysis	115
4.10	Data analysis	118
4.10.1	Descriptive statistics	119
4.10.2	Univariate analysis	120
4.10.3	Multivariate analysis	121
4.11	Biological interpretation of the results	135
4.12	Conclusions	136
4.13	Bibliography	138

CONTENTS

Chapter V: Untargeted metabolomics for plasma biomarker discovery using LC-QTOF-MS

5.1	Introduction.....	153
5.2	Untargeted metabolomics workflow	153
5.2.1	Sampling.....	155
5.2.2	Sample preparation	156
5.2.3	Instrumental analysis.....	157
5.2.4	Data processing.....	158
5.2.5	Data analysis	159
5.2.6	Compound identification	160
5.3	Untargeted metabolomics in nephrology: literature review.....	161
5.4	Study's goal.....	163
5.5	Materials and equipment.....	163
5.5.1	Standards and reagents.....	163
5.5.2	Instrumentation	164
5.5.3	Subjects and sampling	164
5.6	LC-QTOF-MS analytical method.....	165
5.6.1	Optimization of protein precipitation.....	165
5.6.2	Quality control sample preparation.....	169
5.6.3	Instrumental analysis.....	170
5.6.4	Data processing.....	172
5.6.5	Data analysis	175
5.6.6	Metabolite identification.....	184
5.7	Biological interpretation of the results.....	196
5.8	Conclusions	198
5.9	Bibliography.....	199

Chapter VI: LC-QQQ-MS method for potential biomarkers

6.1	Introduction.....	207
6.2	Study's goal.....	208
6.3	Materials and equipment.....	209
6.3.1	Standards and reagents.....	209
6.3.2	Instrumentation	210
6.3.3	Preparation of standards solutions.....	211
6.3.4	Subjects and sampling	212

CONTENTS

6.4	LC-QQQ-MS method development	212
6.4.1	Selection of the chromatographic variables	212
6.4.2	Selection of mass spectrometry detection variables.....	218
6.4.3	Optimization of sample treatment conditions.....	222
6.5	Analytical evaluation of the method	225
6.5.1	Calibration of the method.....	225
6.5.2	Accuracy and precision	226
6.5.3	Stability	228
6.6	Sample analysis.....	230
6.7	Data analysis	232
6.7.1	Descriptive statistics	232
6.7.2	Univariate analysis.....	233
6.7.3	Multivariate analysis.....	234
6.8	Conclusions.....	237
6.9	Bibliography	238

Chapter VII: General conclusions

7.1	General conclusions	243
-----	---------------------------	-----

Annex: Articles published in journal articles

<u>Annex I</u> : LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease	247
<u>Annex II</u> : Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data.....	263



ACRONYMS & ABBREVIATIONS.

ACRONYMS & ABBREVIATIONS

ACEI	Angiotensin-converting enzyme inhibitors
ACN	Acetonitrile
ADMA	Asymmetric dimethylarginine
ADC	Analog-to-digital converters
ADP	Adenosine diphosphate
AGAT	Arginine-glycine amidinotransferase
AJS	Agilent Jet Stream®
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization interface
API	Atmospheric pressure ionization source
APPI	Atmospheric pressure photoionization source
ARA	Angiotensin receptor antagonists
ARG	Arginine
ATP	Adenosine triphosphate
BET	Betaine
BHMT	Betaine-homocysteine methyltransferase
BIL	Bilirubin
BSA	Body surface area
BUN	Blood urea nitrogen
CAKUT	Congenital anomalies of the kidney and the urinary tract
CAN	Acetonitrile
CAPD	Continuous ambulatory peritoneal dialysis
CAPV	Comunidad Autónoma del País Vasco / Autonomous Region of the Basque Country
CBS	Cystathionine beta synthase
CCA	Calcium channel antagonists
CDC	Centers for Disease, Control and Prevention
CE	Capillary electrophoresis
CE-MS	Capillary electrophoresis coupled to mass spectrometry
CAWG	Chemical Analysis Working Group
CEF	Compound exchange format file
CID	Collision induced dissociation
CIS4DEC	<i>cis</i> -4-decenoylcarnitine
CIT	Citrulline
CKD	Chronic kidney disease
CNN	Creatinine
CNN-d ₃	Creatinine-d ₃

ACRONYMS & ABBREVIATIONS

CSV	Comma-separated values file
CTN	Creatine
CTN-d ₃	Creatine-d ₃
CYS	Cysteine
⁵¹ Cr-EDTA	Radioactively tagged ethylenediaminetetra-acetic acid
DC	Direct current
DDAH	Dimethylarginine dimethylaminohydrolase
DMG	Dimethylglycine
dMRM	Dynamic multiple reaction monitoring
DOQI	Dialysis Outcomes Quality Initiative
DTT	Dithiothreitol
ECC	Extracted Compound Chromatogram
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
Eq.	Equation
ERA-EDTA	European Renal Association and European Dialysis and Transplantation Association
ERC	Enfermedad Renal Crónica
ERCT	Enfermedad Renal Crónica Terminal
ESI	Electrospray interface
ESPN	European Society for Paediatric Nephrology
ESRD	End-Stage Renal Disease
ETOH	Ethanol
EUTox	European Uremic Toxin Work Group
FA	Formic acid
Fig.	Figure
Fisher-LDA	Fisher-Linear Discriminant Analysis
FMOC	Fluorenylmethyloxycarbonyl
Frg	Fragmentor voltage
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum
GC	Gas chromatography
GC-MS	Gas chromatography coupled to mass spectrometry
GF	Glomerular filtration
GFR	Glomerular filtration rate
GLY	Glycine
GSH	Reduced glutathione

ACRONYMS & ABBREVIATIONS

GSH- ¹³ C ₂ ¹⁵ N	Glutathione- ¹³ C ₂ ¹⁵ N
GSSG	Oxidized glutathione
γ-GT	Gamma glutamil transpeptidase
HARG	Homoarginine
HCYS	Homocysteine
HILIC	Hydrophilic interaction liquid chromatography
HFBA	Heptafluorobutyric acid
HFBCF	Heptafluorobutyl chloroformate
HMDB	Human metabolome database
iC4	<i>i</i> -butyrylcarnitine
ICR	Ion cyclotron resonance
ID	Identity
IEC-UV	Cation-exchange chromatography with ultraviolet detector
IP-LC-MS	Ion-pairing liquid chromatography coupled to mass spectrometry
IQR	Interquartile range
IS	Internal standard
IT	Ion trap
KEGG	Kyoto Encyclopedia of Genes and Genomes
<i>k</i> NN	<i>k</i> -nearest neighbours
LC	Liquid chromatography
LC-DAD	Liquid chromatography coupled to diode array detector
LC-FLD	Liquid chromatography coupled to fluorescence detector
LC-MS	Liquid chromatography coupled to mass spectrometry
LC-QTOF	Liquid chromatography coupled to quadrupole time-of-flight mass spectrometer
LC-QTOF-MS	Liquid chromatography coupled to quadrupole time-of-flight mass spectrometry
LC-QQQ	Liquid chromatography coupled to triple quadrupole mass spectrometer
LC-QQQ-MS	Liquid chromatography coupled to triple quadrupole mass spectrometry
LC-UV	Liquid chromatography coupled to ultraviolet detector
LDA	Linear discriminant analysis
LLOQ	Lower limit of quantification
LOO	Leave-one-out
LV	Latent variables
MALDI	Matrix-assisted laser desorption ionization interface
MAT	Methionine methyltransferase
MBTFA	N-Methyl-bis(trifluoroacetamide)

ACRONYMS & ABBREVIATIONS

2-ME	2-mercaptoethanol
MEOH	Methanol
MET	Methionine
MCP	Microchannel plate
MDRD	Modification of Diet in Renal Disease
ML-LDA	Maximum likelihood linear discriminant analysis
MPP	Mass Profiler Professional
MRM	Multiple reaction monitoring
MS	Mass spectrometry
MSI	Metabolomics Standards Initiative
MS/MS	Tandem mass spectrometry
MSI	Metabolomics Standards Initiative
MSTFA	N-trimethylsilyl-N-methyl trifluoroacetamide
MTBSTFA	N-methyl-N-tert-butyldimethylsilyltrifluoroacetamide
<i>m/z</i>	Mass-to-charge ratio
nC4	<i>n</i> -butyrylcarnitine
NFPA	Nonafluoropentanoic acid
NMR	Nuclear magnetic resonance
NOS	Nitric oxide sintase
OPA	<i>o</i> -phthaldialdehyde
PC	Principal components
PCA	Principal component analysis
PDFOA	Pentadecafluorooctanoic acid
PFHA	Perfluoroheptanoic acid
PFPCF	2,2,3,3,3-pentafluoropropyl chloroformate
PLP	Pyridoxal phosphate
PLS	Partial least squares regression
PLS-DA	Partial least squares discriminant analysis
Pmarp	Per million age-related population
Pmp	Per million population
PPT	Protein precipitation
Q	Quadrupole
QC	Quality control
QDA	Quadratic discriminant analysis
QTOF	Quadrupole-time-of-flight
QTRAP	Quadrupole-linear ion trap
RAAS	Renin-angiotensin aldosterone system

ACRONYMS & ABBREVIATIONS

RPLC	Reversed-phase liquid chromatography
RF	Radio frequency
ROS	Reactive oxygen species
RP	Resolving power
Rpm	Revolutions per minute
RRT	Renal replacement therapy
RSD	Relative standard deviation
S1P	Sphingosine-1-phosphate
SAH	S-adenosylhomocysteine
SAHH	S-adenosylhomocysteine hydrolase
SAM	S-adenosylmethionine
SCr	Serum creatinine
SDMA	Symmetric dimethylarginine
SDMA-d ₆	Symmetric dimethylarginine
SIM	Selected ion monitoring
SPE	Solid-phase extraction
SPLS-DA	Sparse partial least squares discriminant analysis
SVD	Singular Value Decomposition
⁹⁹ Tc-DTPA	Radioactively tagged diethylene-triamine penta-acetic acid
TDC	Time-to-digital converter
TFA	Trifluoroacetic acid
TIC	Total ion current chromatogram
TIS	Turboionspray
TOF	Time-of-flight mass analyzer
USRDS	The United States Renal Data System
Var	Variance



RESUMEN.

La enfermedad renal crónica (ERC) se define como una condición en la cual los riñones están dañados debido a anomalías funcionales o estructurales y no son capaces de filtrar la sangre adecuadamente, causando la acumulación de desechos en la sangre y en el organismo, lo que puede originar serios problemas de salud.

La ERC en niños se asocia a severas complicaciones de desarrollo, metabólicas o cardiovasculares, que pueden derivar en la necesidad de medidas terapéuticas sustitutivas como la diálisis o el trasplante. A pesar de que los últimos avances en nuevas terapias médicas han permitido un mejor control y pronóstico de estos pacientes, aún no se logra prevenir efectivamente el retraso del crecimiento, la anemia, las alteraciones del metabolismo mineral y otras consecuencias de la enfermedad, alcanzando altas tasas de morbimortalidad cardiovascular en este grupo de población.

El establecimiento de un diagnóstico precoz posibilitaría una intervención terapéutica efectiva para el retardo de su progresión y la aparición de complicaciones. Cabe destacar que a pesar de que la mayoría de los pacientes se encuentran en etapas poco avanzadas de la enfermedad en las que todavía pueden aplicarse medidas terapéuticas efectivas que cambien el curso de la enfermedad, un número reducido de pacientes evoluciona cada año hacia la enfermedad renal crónica terminal (ERCT), en la que se hace necesaria la aplicación de tratamiento de terapia sustitutiva renal (diálisis o trasplante) para la supervivencia.

En la práctica clínica se utiliza la creatinina endógena como biomarcador clásico para evaluar la función renal. Sin embargo, la creatinina se encuentra lejos de ser el biomarcador ideal, ya que su sensibilidad es bastante limitada y revela el daño renal cuando ya se ha producido una importante pérdida de nefronas. Además, existen multitud de inconvenientes como la posibilidad de que se vea afectada por la secreción tubular y por multitud de factores que influyen sobre los niveles de creatinina, como la masa muscular, la dieta, el estado de hidratación o la edad. La búsqueda de nuevos biomarcadores de diagnóstico o evolutivos de la patología y su aplicación en la ERC

pediátrica podría contribuir a anticipar su diagnóstico, mejorar su seguimiento, así como a progresar en la comprensión de los trastornos fisiopatológicos subyacentes.

La metabolómica es una herramienta que tiene como finalidad principal la búsqueda e identificación de biomarcadores mediante el estudio sistemático de perfiles metabólicos en un conjunto de muestras biológicas (fluidos, tejidos, cultivos celulares, etc.), que se centra en el estudio de moléculas pequeñas, típicamente por debajo de los 1500 Da en un sistema biológico. La comparación de los metabolomas entre un grupo control y un grupo testado podría arrojar diferencias en sus perfiles que, asociados a una condición biológica concreta, podría determinar el incremento o disminución de qué metabolitos se asocian a dicha condición. Estos compuestos podrían actuar como biomarcadores de diagnóstico o evolutivos de la enfermedad objeto de estudio.

Se puede diferenciar entre métodos de metabolómica dirigidos y no dirigidos. La metabolómica dirigida o *targeted* se centra en la determinación y cuantificación de metabolitos conocidos, que se sospecha puedan hallarse alterados en una determinada enfermedad, mientras que la metabolómica no dirigida o *untargeted* posibilita analizar y comparar tantos metabolitos como sea posible sin ningún tipo de sesgo e identificar estas entidades *a posteriori*. Dado que ambas propuestas podrían resultar de utilidad a la hora de buscar nuevos biomarcadores, se ha trabajado en ambos planteamientos.

Para ello se han utilizado muestras de plasma de pacientes pediátricos con ERC y controles del País Vasco, recogidos por el Servicio de Nefrología Pediátrica del Hospital de Cruces (Barakaldo, España), que han sido alicuotadas para permitir su utilización en los distintos planteamientos.

Con el objetivo de encontrar nuevos biomarcadores de esta enfermedad, el presente estudio se ha estructurado alrededor de tres objetivos operacionales:

En primer lugar, se ha desarrollado y validado un método de metabolómica dirigido por cromatografía de líquidos acoplada a espectrometría de masas con analizador de tipo

cuadrupolo tiempo de vuelo (LC-QTOF) basado en el estudio de 16 aminoácidos, derivados aminoacídicos y compuestos relacionados pertenecientes a las rutas metabólicas arginina-creatina, metilación de la arginina y ciclo de la urea, que se sospechaba podían verse alterados en pacientes pediátricos con ERC. Teniendo en cuenta que los aminoácidos y derivados aminoacídicos son compuestos pequeños y polares, ha sido necesario el uso de ácido perfluoroheptanoico como par iónico para mejorar su retención en cromatografía líquida. Además, debido a la naturaleza de algunos de estos compuestos, se ha utilizado dithiothreitol como agente reductor, a fin de permitir la reducción de los aminothioles para su cuantificación total.

Tras llevar a cabo el análisis de las muestras de plasma pertenecientes a pacientes pediátricos controles y enfermos con ERC, se han realizado tanto un análisis estadístico univariante como multivariante a fin de interpretar los resultados obtenidos.

El análisis univariante permitió encontrar diferencias significativas entre ambos grupos para los siguientes compuestos: glicina, citrulina, creatinina, dimetilarginina asimétrica, dimetilarginina simétrica y dimetilglicina.

Por su parte, el análisis multivariante mostraba un incremento en los niveles plasmáticos de los pacientes con ERC para los siguientes metabolitos: dimetilarginina simétrica, S-adenosilhomocisteína, creatinina, S-adenosilmetionina, citrulina, dimetilarginina asimétrica, glutatión, dimetilglicina y glicina. Tras llevar a cabo una reducción de variables, se proponen citrulina, S-adenosilmetionina y dimetilarginina simétrica como biomarcadores potenciales que, junto a la creatinina (cuya capacidad de clasificación correcta de las muestras es del 71 %), podrían ser utilizadas como biomarcadores potenciales para mejorar el diagnóstico de los estadios primarios de la enfermedad, hasta alcanzar el 89 % de clasificación acertada. Señalar también que estos cuatro metabolitos permiten incremento en la predicción del grupo al que pertenecen en un 10 % en comparación con el uso exclusivo de la creatinina para el diagnóstico de la ERC sin tener en cuenta el grado de enfermedad que padecen. Además, se ha hallado una gradación en la concentración de estos compuestos que se corresponde con el grado de la enfermedad.

Por todo ello, se considera razonable incluir estos 3 metabolitos en una nueva ecuación que pueda ser utilizada en las pruebas de cribado llevadas a cabo por médicos de familia, a fin de aproximar el diagnóstico de la enfermedad que habitualmente llevan a cabo los profesionales de la nefrología, permitiendo prever el grado de ERC que sufre antes de derivarlo al especialista.

En segundo lugar, se ha desarrollado un método de metabolómica no dirigida en LC-QTOF, con el objetivo de comparar los perfiles metabolómicos sin ningún tipo de sesgo y hallar aquellos metabolitos que pudieran hallarse alterados en los niños con ERC. En este caso, el flujo de trabajo ha consistido en la adquisición de los datos, el preprocesamiento de la señal mediante la eliminación del ruido, detección y alineamiento de picos y deconvolución, seguido de un análisis estadístico y quimiométrico de los datos. Finalmente, las entidades obtenidas fueron identificadas utilizando distintas bases de datos y su identidad fue confirmada mediante la comparación de los tiempos de retención, espectro de masas y espectro de MS/MS de las entidades obtenidas en el plasma con estándares analíticos.

Teniendo en cuenta que las distintas estrategias quimiométricas pueden afectar al número de entidades obtenidas, se emplearon dos flujos de trabajo diferentes utilizando como herramientas Matlab (Mathworks) y Mass Profiler Professional (Agilent Technologies), y se consideraron únicamente aquellas entidades coincidentes en ambas estrategias. De este modo, se obtuvieron 5 entidades que se encontraban significativamente elevadas o disminuidas en los pacientes con ERC. La utilización de estas entidades permitiría la discriminación de los pacientes con ERC y controles en un 96 %, mientras que la clasificación de pacientes en estadios tempranos de la enfermedad se produce correctamente en el 97 % de los casos.

Dado que estas 5 entidades resultaban prometedoras de cara a mejorar la clasificación de las muestras con ERC y controles, se procedió a su identificación mediante el uso de bases de datos de espectrometría de masas, la adquisición de espectros MS/MS de estas entidades y el uso de estándares analíticos. Cuatro de estas cinco entidades fueron

identificadas con el máximo nivel de confianza, de acuerdo a los niveles establecidos por la *Metabolomics Standards Initiative*, como: *n*-butirilcarnitina, *cis*-4-decenoilcarnitina, bilirrubina y esfingosina-1-fosfato.

Para finalizar el presente estudio, se ha desarrollado un método de cuantificación por cromatografía de líquidos acoplada a espectrometría de masas con analizador de tipo triple cuadrupolo (LC-QQQ), a fin de reunir las entidades más relevantes obtenidas tanto mediante el estudio de metabolómica dirigida como en el de metabolómica no dirigida. Tras optimizar y validar el método LC-QQQ para la cuantificación de citrulina, creatinina, dimetilarginina simétrica, S-adenosilmetionina, *n*-butirilcarnitina, *cis*-4-decenoilcarnitina, esfingosina-1-fosfato y bilirrubina, éste se aplicó a muestras reales de pacientes pediátricos con ERC y controles.

Por un lado, los resultados obtenidos para los aminoácidos y compuestos relacionados se correspondían con los obtenidos en el primer método dirigido desarrollado, hallándose los niveles de concentración de citrulina, creatinina, dimetilarginina simétrica y S-adenosilmetionina significativamente aumentados en los niños con ERC. Por otro lado, en lo que a la validación de los resultados de metabolómica no dirigida se refiere, *n*-butirilcarnitina, *cis*-4-decenoilcarnitina y esfingosina-1-fosfato presentaban niveles de concentración significativamente superiores en los pacientes pediátricos con ERC. Sin embargo, a pesar de que los niveles de bilirrubina se encontraban disminuidos significativamente en el estudio no dirigido llevado a cabo previamente, este último método LC-QQQ no demostró una disminución significativa para la bilirrubina.

Finalmente, en lo que concierne a la separación de las muestras en función de las concentraciones obtenidas para los distintos metabolitos, es posible diferenciar los grupos de paciente control y enfermo en las representaciones obtenidas a partir de la realización de análisis de componentes principales. Es por ello que se considera recomendable el estudio de una población más amplia de cara a obtener la significación de la bilirrubina como biomarcador potencial y evaluar la capacidad de clasificación de las muestras en

controles y enfermos de este nuevo método, a fin de que pueda utilizarse en la práctica clínica.



SUMMARY.

Chronic kidney disease (CKD) is defined as a condition in which kidneys are damaged due to functional or structural abnormalities and are not capable of filtering blood adequately, causing the accumulation of waste products in blood and in the organism, originating serious health problems.

Paediatric CKD is associated to severe development, metabolic or cardiovascular complications that may lead to the need for renal replacement therapeutic measures, such as dialysis or transplant. Although the latest therapeutic breakthroughs have allowed a better control and prognosis of these patients, it is not yet possible to effectively prevent growth retardation, anaemia, alterations in mineral metabolism and other consequences of the disease, reaching high rates of cardiovascular morbimortality in this population group.

The establishment of an early diagnosis would allow an efficient therapeutic intervention to delay the progression of CKD and the occurrence of complications. Indeed, even though the majority of the patients are in early stages of the disease in which effective therapeutic measures might be adopted to be able to change the course of the disease, a reduced number of patients evolve every year toward end-stage renal disease (ESRD), in which renal replacement therapy becomes essential (either dialysis or transplant) for survival.

In clinical practice, endogenous creatinine is the classical biomarker used to assess renal function. However, creatinine is far from being the ideal biomarker, since its sensitivity is limited and reveals renal damage when an important loss of nephrons have already occurred. Moreover, numerous drawbacks have been identified, such as the possibility of being affected by tubular secretion or multiple factors that influence creatinine levels like muscular mass, diet, hydration status or age.

The discovery of new diagnostic biomarkers or markers of evolution of the pathology and their application to paediatric CKD could contribute to foresee its diagnosis, make better follow-up as well as to improve the understanding of underlying pathophysiological disorders.

Metabolomics is a powerful tool focused on identifying metabolite biomarkers through a systematic study of metabolic profiles in a set of biological samples (biofluids, tissues, cell cultures, etc.), aimed at small molecules, typically under 1500 Da in a biological system. Comparison of metabolomes of control and test groups could show differences in their profiles related to a specific biological condition, and the increased or decreased metabolites associated to that condition could be determined. These compounds could act as diagnostic or evolutionary biomarkers of the disease under study.

Metabolomics methods might be targeted or untargeted. Targeted metabolomics focuses on the determination and quantification of known metabolites, which are suspicious of being altered in a specific disease, whereas untargeted metabolomics makes possible the analysis and comparison of as many metabolites as possible without any bias, followed by subsequent identification of the entities. Regarding that both methodologies could be of interest aimed at finding new biomarkers, working in targeted and untargeted metabolomic approaches was decided.

For that purpose, plasma samples from paediatric patients with CKD and controls from the Basque Country have been used, collected by the Paediatric Nephrology Service at Cruces University Hospital (Barakaldo, Spain), which had being aliquoted to enable their utilization in different approaches.

With the aim of obtaining new biomarkers for CKD, this study has been structured around three operational objectives:

First of all, a targeted metabolomics method has been developed and validated using liquid chromatography coupled to mass spectrometry with quadrupole-time-of-flight (LC-QTOF) analyzer based on the study of 16 amino acids, amino acid derivative and related compounds belonging to arginine-creatine, arginine methylation and urea cycle metabolic pathways, which were suspicious of being altered in pediatrics suffering from CKD. Regarding that amino acids and amino acid derivatives are small and polar compounds, the use of perfluoroheptanoic acid as an ion-pairing reagent has been necessary to improve their retention in liquid chromatography. In addition, concerning the nature of

some of these compounds, dithiothreitol has been used as reducing agent, aimed at enabling the reduction of aminothiols for their total quantification.

Analysis of plasma samples corresponding to control and CKD paediatrics was carried out with this targeted method and univariate and multivariate statistical analyses were performed in order to interpret the results obtained.

Univariate analysis enabled discovering significant differences between groups for the following compounds: glycine, citrulline, creatinine, asymmetric dimethylarginine, symmetric dimethylarginine and dimethylglycine.

As for multivariate analysis, it showed an increase of plasmatic levels of the following metabolites in patients with CKD: symmetric dimethylarginine, S-adenosylhomocysteine, creatinine, S-adenosylmethionine, citrulline, asymmetric dimethylarginine, glutathione, dimethylglycine and glycine. After variable reduction, citrulline, S-adenosylmethionine and symmetric dimethylarginine were proposed as potential biomarkers that together with creatinine (with a capacity for correct sample classification rate of 71 %), could be used as potential biomarkers to improve the diagnosis of primary stages of the disease, reaching 89 % of right classification. It should be noted that these 4 metabolites also enable increasing 10 % the prediction of their membership in comparison with the exclusive use of creatinine for CKD diagnosis, regardless of the degree of CKD suffered.

Besides, a gradation in the concentration of these compounds have been found which corresponds to the degree of the disease. For all these reasons, it would be reasonable including these 3 metabolites in a new equation which could be used in the screening tests carried out by general practitioners, with the aim of approximating the diagnosis of the disease, which is usually carried out by nephrologists, to foresee the CKD degree suffered before referring patients to specialist.

Secondly, an untargeted metabolomics method has been developed in LC-QTOF equipment with the objective of comparing metabolomic profiles without any bias and finding metabolites that might be altered in paediatrics with CKD. In this case, workflow

consisted in data acquisition, signal preprocessing by means of background elimination, detection and alignment, followed by statistical and chemometrical analyses of data. Finally, features obtained were identified using different databases and their identity was confirmed by means of comparison of retention time, exact mass and MS/MS spectra of the entities obtained in plasma with analytical standards.

Concerning that different chemometric approaches used could affect the number of entities obtained, two different workflows were adopted using Matlab (Mathworks) and Mass Profiler Professional (Agilent Technologies) tools, and only the entities found to be coincident in both strategies were considered. This way, 5 entities were found to be significantly increased or decreased in patients with CKD. The use of these entities would enable discrimination of patients with CKD and controls with a performance of 96 %, whereas classification of patients suffering from early CKD would be successful in 97 % of the cases.

Regarding that these 5 entities would be promising to improve classification of CKD and control samples, we proceeded to their identification by means of mass spectrometry databases, MS/MS spectra acquisition and the use of analytical standards. Four out of five of these entities were identified with the maximum confidence level, according to the standards set by Metabolomics Standards Initiative, as: *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine, bilirubin and sphingosine-1-phosphate.

Finally, a quantification method has been developed by means of liquid chromatography coupled to mass spectrometry with triple-quadrupole (LC-QQQ) mass spectrometer, aimed at gathering the most relevant metabolites obtained from previous targeted and untargeted metabolomics studies. After optimizing and validating LC-QQQ method for citrulline, creatinine, symmetric dimethylarginine, S-adenosylmethionine, *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine, sphingosine-1-phosphate and bilirubin, it was applied to real samples of paediatrics with CKD and controls.

On the one hand, the results obtained for the amino acids and related compounds were in accordance to those obtained in the first targeted method developed, being revealed

citrulline, creatinine, symmetric dimethylarginine and S-adenosylmethionine concentration levels increased in paediatrics with CKD. On the other hand, regarding validation of the biomarkers obtained from untargeted metabolomics, *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine and sphingosine-1-phosphate showed to be significantly increased in paediatrics with CKD in comparison with control patients. However, although bilirubin levels were decreased in the untargeted metabolomics study carried out previously, this LC-QQQ method did not show a significant decrease for bilirubin.

Finally, concerning separation of samples according to the metabolite concentrations, it is possible to differentiate control and CKD patient groups in the representations obtained from principal component analysis algorithm. For all these reasons, it is considered advisable studying a larger population aimed at verifying the significance of bilirubin as potential biomarker, as well as evaluating the classification capacity of samples into control and ill patients with this new method in order to be usable in the clinical practice.



CHAPTER I.

INTRODUCTION

- 1.1. Chronic kidney disease. Definition**
- 1.2. Etiopathology of CKD**
- 1.3. CKD in adults and in paediatrics**
- 1.4. CKD degree**
- 1.5. Epidemiology of CKD**
- 1.6. Risk factors for CKD progression and complications of CKD**
- 1.7. CKD treatment**
- 1.8. Diagnosis and evaluation of CKD**
- 1.9. Biomarkers. Compounds of interest for the early detection and intervention on the disease**
- 1.10. Bibliography**

1.1. CHRONIC KIDNEY DISEASE. DEFINITION

Chronic kidney disease (CKD) is increasingly recognized as a global public health problem and even World Kidney Day was declared in 2006. Moreover, the acceptance of CKD as a public health pathological condition has evolved more homogeneous definitions and classifications of CKD¹.

The *Centers for Disease, Control and Prevention* (CDC) describes CKD as a condition in which the kidneys are damaged and are not able to filter the blood as healthy kidneys would do, causing waste accumulation in blood and in the organism, thus potentially originating health problems².

In addition, the *National Kidney Foundation*, through the *Dialysis Outcomes Quality Initiative* (DOQI), publishes guidelines about CKD, which have had a significant impact in the care and evolution of patients undergoing dialysis. These guidelines itemize the two criteria from which at least one must be met to consider that a renal affection must be defined as CKD (see Table 1.1).

Table 1.1. Criteria to consider that a patient suffers from CKD. Extracted from *K/DOQI Clinical Practice Guidelines For Chronic Kidney Disease: Evaluation, Classification and Stratification*³.

CRITERIA TO DEFINE CKD
<p>1. Kidney damage for ≥ 3 months, defined by structural or functional abnormalities of the kidneys, with or without decreased GFR, which can be manifested by either:</p> <ul style="list-style-type: none"> - Pathological abnormalities, or - Markers of kidney damage, including abnormalities in the composition of blood or urine, or abnormalities in diagnostic imaging tests
<p>2. GFR <60 mL/min/1.73m² for ≥ 3 months, with or without kidney damage</p>

1.2. ETIOPATHOLOGY OF CKD

The etiopathology of the disease can be better understood with some basic knowledge about the renal system.

The renal system, also known as urinary system, consists of two kidneys, two ureters, bladder and urethra. Kidneys are two bean-shaped reddish-brownish organs located in retroperitoneal conjunctive tissue, in the rear of the abdomen (see Fig. 1.1). Kidney size is usually about 10 cm long, 5 cm wide and 2.5 cm deep for adults. Despite having a

similar size and shape, left kidney is slightly longer and thinner than the right one and is situated closer from the midline^{4,5}.

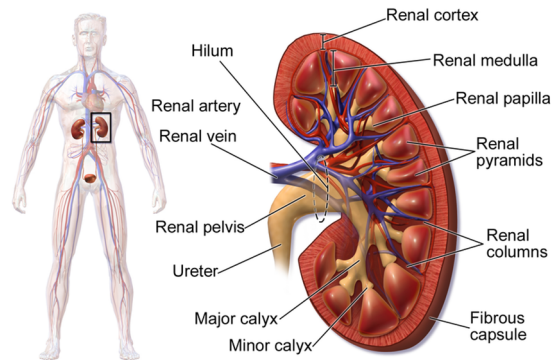


Fig. 1.1. Kidney location and anatomy, adapted from Blaus et al⁶.

Kidneys are responsible for almost all the activity in the urinary system, but also regulate many other important body functions. Some of the core functions of the kidneys include^{7,8}:

- ✓ Regulation of ionic composition of blood
- ✓ Blood pH adjustment
- ✓ Blood pressure regulation
- ✓ Preservation of osmolarity
- ✓ Hormone production
- ✓ Regulation of blood glucose levels
- ✓ Excretion of waste products from metabolic reactions in the organism

Most of these kidney functions are managed by nephrons, the functional units of kidneys. Nephrons consist of a renal corpuscle, made of glomerulus and Bowman capsule, and a renal tubule, comprising a proximal tubule, Henle’s loop and a distal tubule⁷, as showed in Fig. 1.2.

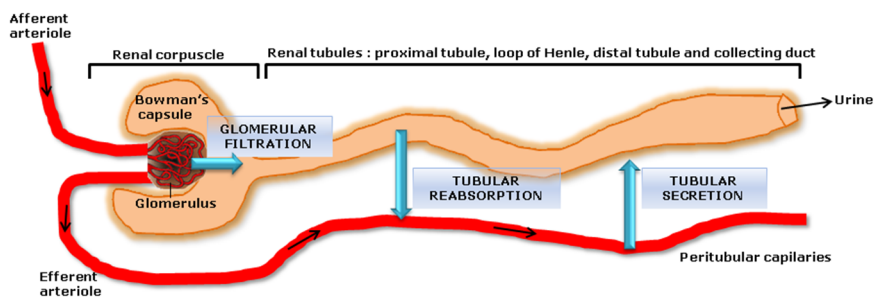


Fig. 1.2. Graphic representation of a nephron.

Blood is filtered through each nephron's Bowman's capsule and then, the filtrated liquid is conducted to renal tubules. Tubular reabsorption mechanisms in the tubules return the majority of water and some solutes back to the circulatory system. In addition, secretion processes allow the excretion of unwanted substances from the bloodstream. Finally, the filtrate is drained into a collector tubule, and the collector tubules join together into new conducts^{7,8}.

Glomerular filtration (GF) is the quantity of filtrate formed in all the renal corpuscles from both kidneys per minute. Preserving a constant GF is essential to keep homeostasis. If GF is too high, necessary substances will go through renal tubules too fast to be reabsorbed and therefore they could be lost in urine, whereas if GF is too low, almost all the filtrate will be absorbed back and some waste products could not be excreted adequately. For that reason, kidneys themselves help keeping kidney blood flow and GF constant by means of myogenic and feedback mechanisms. More than the 99 % of the glomerular filtrate is returned to the bloodstream and therefore, only 1-2 L of urine are excreted per day⁷.

CKD can begin indiscriminately in glomeruli, tubuli or renal vessels. In accordance with the affected area of the kidneys, different causes for CKD are reported in Table 1.2.

Table 1.2. Etiopathogenesis of CKD⁹.

Glomerular diseases	<ul style="list-style-type: none"> ➢ Diabetes ➢ Hypertension ➢ Systemic and renal infections ➢ Focal and segmental glomerulonephritis ➢ Immune deposits in the capillary wall (lupus nephritis, membranoproliferative glomerulonephritis), accumulation of IgA complexes in the glomeruli (IgA nephropathy) and others ➢ Vasculitic glomerulonephritis
Tubular diseases	<ul style="list-style-type: none"> ➢ Idiopathic ➢ Genetic ➢ Chemical action of drugs and toxics ➢ Infection and inflammation of tubulointerstitium ➢ Increased intratubular pressure due to mechanical stress and related to obstruction of lower urinary tract ➢ Transplant rejection due to immune response
Renovascular diseases	<ul style="list-style-type: none"> ➢ Hypertension ➢ Renal oxidative stress, endothelial function and inflammation initiated by atherosclerosis ➢ Hypoperfusion and ischemic scenarios (for instance, due to stenosis)

Glomerulopathies, alterations and diseases affecting glomerular structure, could cause CKD⁹. Glomerulonephritis, a type of primary glomerulopathy with an inflammatory component, is in addition to diabetes and hypertension one of the most common causes for CKD related to glomerular diseases^{10,11}.

On the other hand, renal function could also be damaged due to dysfunctional tubular reabsorption and secretion processes, activation of tubular cells by means of inflammatory mediators, progressive tubular loss and the healing process of the tissue. Moreover, other renal structures such as glomeruli could be damaged afterwards³.

Finally, renovascular alterations cover the damage originated from atherosclerosis, hypertension, hypoperfusion and ischaemia. For instance, atherosclerosis leads to an oxidative stress process, which is accompanied by endothelial dysfunction and inflammation, thus conducting to fibrosis and glomerular filtration reduction⁹.

Whatever the origin of the damage, the progression to CKD presents histological and functional changes affecting all renal structures. These changes lead to progressive and generalized fibrosis, nephron loss and its replacement with scar tissue. As the disease progresses, structural damage affects the function of a larger number of nephrons. At first, less injured nephrons adapt themselves to balance dysfunctional nephrons. That is the reason why glomerular filtration does not decrease until approximately 60% of the nephrons are affected and functionality is compromised. At this point, functional nephrons are not able to clear blood waste products by filtration and elimination, and uraemia occurs^{9,12}.

1.3. CKD IN ADULTS AND IN PAEDIATRICS

The basic physiopathologic mechanisms of CKD are shared in adults and in paediatrics. However, the etiopathology or the primary causes of CKD differs between adults and paediatrics, as showed in Fig. 1.3. For instance, in adults the most prevalent primary causes (72 %) of CKD are hypertension and diabetes, whereas in paediatrics congenital anomalies of the kidney and the urinary tract (CAKUT) and primary and secondary glomerular diseases (e.g. lupus nephritis, glomerulonephritis, atypical haemolytic uremic syndrome, Alport syndrome) account for approximately 56 %¹².

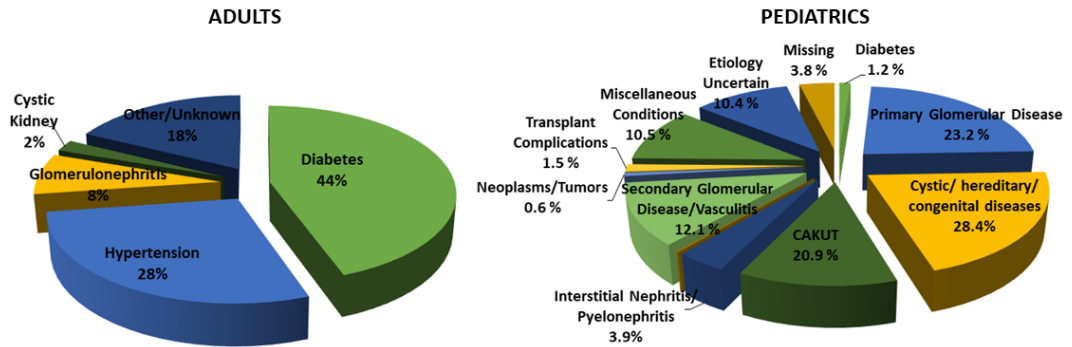


Fig. 1.3. Primary cause of ESRD differs between adults and children, as showed in this graph obtained from the 2016 USRDS report¹².

Moreover, congenital anomalies are the most common underlying etiology of paediatric End-Stage Renal Disease (ESRD) population, those suffering from the most severe stage of CKD. As a consequence, paediatrics face unique problems in preparation for transplant¹³. In addition, paediatrics suffer from specific complications not present in adults, such as growth retardation^{14,15}.


For all these reasons, CKD in paediatrics has sometimes been recognized as an independent nosologic entity¹⁶.

1.4. CKD DEGREE

CKD can be classified according to severity, diagnosis, treatment and prognosis both in adults and in paediatrics. However, the most accepted classification worldwide is the one shown in Table 1.3, which considers not only severity but also some other markers of kidney damage.

In clinical practice, glomerular filtration rate (GFR), glomerular filtration (GF) given as a relation to body surface area (BSA) by scaling the value to 1.73 m², is more commonly used than GF. GFR is on average 100-130 mL/min/1.73 m² in normal children and young adults aged 2 to 16 years old. Despite the global implementation of this CKD classification criterium, it has to be noted that it is not used on children under 2-year age, as GFR increases from birth and steady values are not reached until that age¹⁷.

Table 1.3. Classification of CKD. Adapted from Levey *et al.* and based on the K/DOQI Clinical Practice Guidelines For Chronic Kidney Disease classification^{3,18}.

Stage	Description	GFR (mL/min/1.73 m ²)	Related terms	
Mild 	1	Kidney damage with normal or ↑ GFR	≥ 90	Albuminuria, proteinuria, haematuria
	2	Kidney damage with mild ↓ GFR	60-89	Albuminuria, proteinuria, haematuria
	3	Moderate ↓ GFR	30-59	Chronic renal insufficiency, early renal insufficiency
	4	Severe ↓ GFR	15-29	Chronic renal insufficiency, late renal insufficiency, pre-ESRD
Severe	5	Kidney failure	< 15 (or dialysis)	Renal failure, uraemia, ESRD

CKD stage 1 is often considered a silent stage as in most cases it is characterized by the absence of any symptoms. On the contrary, CKD stage 2 may show an increase of parathyroid hormone and a decrease of renal calcium reabsorption, in addition to a decrease of GFR. Similarly, left ventricular hypertrophy and anaemia secondary to erythropoietin deficiency have been observed in patients in CKD stage 3. Finally, patients with CKD stage 4 often show increased serum triglycerides, hyperphosphatemia, hyperkalemia, metabolic acidosis, fatigue, nausea, anorexia and bone pain, which are subsequently followed by renal failure and severe symptoms of uraemia in CKD stage 5⁹. During CKD stage 5, End-Stage Renal Disease (ESRD) occurs. ESRD is the last and most severe stage of CKD, defined by the presence of uraemia and the need for chronic renal replacement therapy (renal transplantation, peritoneal dialysis or haemodialysis) to reverse the symptoms of uraemia and prolong life¹⁹.

1.5. EPIDEMIOLOGY OF CKD

Epidemiology in adults show that in North America, up to 11% of the population (19 million people) may suffer from CKD²⁰. Similarly, data from Europe, Japan and Australia depict a prevalence of CKD of 6-16 %^{21,22}. In addition, incidence rate for patients with CKD receiving renal replacement therapy is estimated to be between 92 and 185 per million population (pmp) depending on the European country²³. Besides, the last epidemiology report published by the Basque Health System (Osakidetza) reveals that incidence of CKD

patients within the Autonomous Community of the Basque Country (CAPV, Spain) keeps stable around 100 pmp, whereas the total amount of patients with CKD is growing and reached in 2011 a prevalence of 1071 pmp²⁴.

Regarding epidemiology of the disease in paediatric population, information is limited as CKD is frequently asymptomatic and therefore under-diagnosed and underreported²⁵. Moreover, there is still insufficient information on the early stages of paediatric CKD and the available epidemiological data is generally related to ESRD registries²⁶.

The Italkid Project is a prospective, population-based registry for assessing the epidemiology of paediatric CKD. The registry started in 1990 and its inclusion criteria are the following: a creatinine clearance $<75 \text{ mL/min/1.73m}^2$, and an age limit of 20 years at the time of registration. According to Italkid Project, between 1995 and 2000 the average incidence was 12.1 cases per million age-related population (pmarp) and at 1st January 2001 prevalence was 74.7 pmarp²⁷.

In addition, according to the registries collected by the *European Society for Paediatric Nephrology* (ESPN) and the *European Renal Association and European Dialysis and Transplantation Association* (ERA-EDTA) between 2009 and 2011, prevalence of children receiving renal replacement therapy (RRT) was 27.9 pmarp in patients aged 0–14 years, taking into account 37 European countries, whereas the incidence of children in RRT was 5.5 pmarp. Furthermore, mortality in paediatric patients being treated with RRT was 55-fold higher than in general paediatric population²⁸. Prevalence and incidence data of CKD in Europe and in the Autonomous Region of the Basque Country for adults and paediatrics is similar according to ERA-EDTA registries²⁹, as showed in Table 1.4. Data of cases pmarp is facilitated for both adults and paediatrics. In addition, the proportion of CKD patients according to the age range is expressed as a percentage (%) and for paediatrics total amount of cases is also showed (n).

Table 1.4. Incidence and prevalence data of RRT in children and in adult population with CKD by ages in Europe and in Basque Country in 2011²⁹.

	0-19 years		20-44 years		45-65 years		65-74 years		+ 75 years	
	pmarp	%	pmarp	%	pmarp	%	Pmarp	%	pmarp	%
Incidence										
Europe†	5.3(*)	1	46.0	12	155.7	31	369.5	25	476.7	30
Basque C.	10.4	2	42.1	13	107.3	27	356.1	30	296.5	28
Prevalence										
Europe†	34.6(**)	1	484.6	18	1450.4	40	2308.8	22	2283.4	20
Basque C.	83.5	1	500.2	16	1567.8	40	2805.1	24	2049.0	19
	0-4 years n(pmarp)		5-9 years n(pmarp)		10-14 years n(pmarp)		15-19 years n(pmarp)			
Incidence (*)										
Europe†	150(7.9)		71(3.7)		139(7.4)		266(13.4)			
Prevalence (**)										
Europe†	206(20.9)		321(32.3)		565(57.4)		1019(98.9)			

†It has to be noted that the availability of data is limited and does not cover all the European Countries.

Similarly, according to the last epidemiology report from the Basque Health System (Osakidetza), during 2011 fifty paediatric patients followed RRT in Paediatric Nephrology Service at Cruces Hospital, being 5 of them incident on dialysis that year. In addition, in the Autonomous Region of the Basque Country (CAPV), 115 renal transplants were carried out in paediatric patients from 1993 to the end of 2011. Moreover, 50 % of the transplanted patients received their first transplant after a period of 8.5 months²⁴.

1.6. RISK FACTORS FOR CKD PROGRESSION AND COMPLICATIONS OF CKD

A number of factors are related to increased risk of CKD both in adults and in paediatrics. Some of these risk factors have also been considered complications of CKD, because they worsen with the progression of the disease. As the line between risk factors and complications of CKD is still too subtle and the role of some of these manifestations is still unclear, both risk factors and complications of CKD are summarized together in Table 1.5.

Table 1.5. Risk factors and complications of CKD.

RISK FACTORS AND COMPLICATIONS	
<p><u>Non-modifiable risk factors:</u></p> <ul style="list-style-type: none"> ✓ Low birth weight³⁰ ✓ African American ethnicity³⁰⁻³² ✓ Puberty age³³ ✓ Genetic factors (polymorphisms in the angiotensin converting enzyme gene, angiotensinogen, pro-inflammatory molecules, pro-fibrotic molecules or angiogenic factors)³³ 	<p><u>Modifiable risk factors:</u></p> <ul style="list-style-type: none"> ✓ Hypertension (prevalent in paediatrics with CKD and extremely uncommon in healthy paediatrics)³⁴ ✓ Cardiovascular disease^{15,35} ✓ Proteinuria^{33,36,37} ✓ Anaemia³³ ✓ Morbid obesity^{38,39} ✓ Hyperphosphatemia⁴⁰ ✓ Smoking³³
<p><u>Complications:</u></p> <ul style="list-style-type: none"> ✓ Increased cardiovascular risk^{15,35,41} ✓ Anaemia¹⁵ ✓ Proteinuria^{33,36,37} ✓ Lipoprotein abnormalities^{15,33,42} ✓ Growth retardation in paediatrics^{14,15,43,44} ✓ Malnutrition⁴⁵ ✓ Limitations in the emotional, social and functional domains of adult life¹³ ✓ Vitamin D deficiency^{46,47} 	<p><u>Unknown whether it is a primary or secondary event:</u></p> <ul style="list-style-type: none"> ✓ Elevated levels of uric acid³³ ✓ Elevated markers of inflammation (IL-1, IL-6, TNF-α, IFN-γ and C-reactive protein)^{48,49}

All these complications lead to a 30 times higher mortality rate in paediatrics suffering from ESRD in comparison with healthy children²⁵. For that reason, a multidisciplinary approach is required to assess CKD related risk factors and complications in paediatrics suffering from this disease, and reduce their morbidity and mortality.

1.7. CKD TREATMENT

Prior to the 1950's patients with failing renal function were doomed to death and research and technology developed for renal replacement therapy (RRT) involved a medical breakthrough for patients' prospective. In 1933, the first human kidney allotransplant was achieved by Yurii Voronoy using an anoxic cadaver kidney⁵⁰. Soon after, in 1960, due to the invention of the Scribner-Quinton Shunt by Dr. Belding Scribner, long-term haemodialysis was made possible for patients suffering from CKD at the University of Washington Hospital in Seattle (USA)⁵¹. Likewise, in 1959, Richard Ruben used peritoneal

dialysis (PD) successfully on a patient with CKD⁵². Later, automated PD equipment and a prototype for home PD equipment were developed by Henry Tenckhoff and his team⁵³.

Currently there are no specific cures for CKD and renal transplantation is limited by organ scarcity. Therefore, present efforts are focused on the prevention of the progression of CKD⁵⁴. CKD treatment needs for a multidisciplinary approach, based on nutrition, pharmaceutical therapy and depending on the degree of CKD, renal replacement therapy.

Concerning pharmacotherapy, decreasing blood pressure alone has proved to be effective reducing proteinuria to some extent^{55,56} and for that reason, antihypertensive agents may have significant abilities to reduce proteinuria^{56,57}. Renin-angiotensin aldosterone system (RAAS) inhibition by means of angiotensin-converting enzyme inhibitors (ACEI) and angiotensin receptor antagonists (ARA) is considered the most effective therapy to slow CKD progression and reduce proteinuria in adults^{54,58,59}. However, there is limited data on the effects of ACEI and ARA in paediatric patients and the results are not conclusive⁶⁰⁻⁶². In addition, calcium channel antagonists (CCA) are a family of compounds which are used as part of multiple drug therapies with the aim of normalizing blood pressure in patients with CKD in which one drug is not enough to achieve normotension^{63,64}.

Lipid-lowering therapy is another frequent approach for the prevention of cardiovascular morbidity and mortality in adults with CKD. Besides, statin therapy may slow renal disease progression, decrease anti-inflammatory properties and oxidative stress and limit proteinuria in adults^{65,66}.

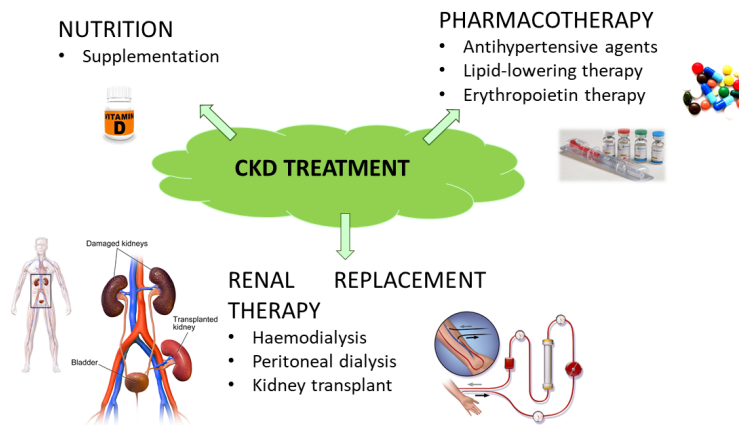


Fig. 1.4. Overview of the multidisciplinary treatment of CKD.

Furthermore, erythropoietin therapy used to treat mild to moderate anaemia in adults with CKD may slow the progression of kidney disease⁸. However, for the moment there is no data available regarding the effectiveness of erythropoietin therapy to slow the GFR decrease in children with CKD⁶⁷.

Concerning the nutritional approach, for decades dietary protein restriction has been hypothesized to diminish renal injury. However, nowadays there is no evidence of any correlation between low-protein diet and progression of CKD in children⁶⁸. Identifying children with inadequate calorie, protein and trace element intake can be challenging but it is necessary for an early nutritional intervention by means of supplementation⁶⁹.

The progressive and irreversible deterioration of the renal function in CKD often results in the implementation of artificial support of the functions of impaired kidneys. This artificial support is known as renal replacement therapy (RRT). Different modalities of RRT are available and recommended to CKD patients depending upon their own characteristics: dialysis and renal transplantation⁷⁰.

Dialysis is based on the use of a semi-permeable membrane that allows diffusive clearance of blood. Blood may be passed through artificial membranes or tubing (haemofiltration or haemodialysis) or dialysate may be instilled contiguous to the peritoneal membrane (peritoneal dialysis). Either peritoneal dialysis or haemodialysis may be used in intermittent or continuous therapies. Despite the fact that there is not any definite level of serum creatinine concentration that mandates starting with dialysis, it is generally accepted that RRT should be considered when GFR declines to a value between 9 and 14 mL/min/1.73m² body surface area^{71,72}.

Regarding renal transplantation, it replaces native renal function completely, although the availability of organs is the main limitation for the widespread use of this technique⁶⁹. Even though renal transplantation is widely recognized as the RRT of choice for paediatrics with ESRD, allograft rejection may occur, which is an inflammatory reaction associated with immune response on the graft. For that reason, transplantation tolerance needs to be carefully considered and a complementary immunosuppressive pharmacotherapeutic approach needs to be fulfilled⁷³.

1.8. DIAGNOSIS AND EVALUATION OF CKD

Traces of the interest for medical conditions related to kidneys have been found since ancient times, for instance, in some cuneiform clay tablets from Mesopotamia⁷⁴. In addition, a few causes of CKD have been familiar for more than 500 years, such as nephrotic syndrome, described by Cornelius Roelans in *De aegri tudinibus infantium* (1484). During the 19th century, even though pathology was restricted to macroscopy, the German pathologist Friedrich Theodor von Frerichs (1819-1885) correlated clinical symptoms of kidney diseases and chemical disorders in blood and urine⁷⁵. Later, during the same century, the use of autopsies as well as the discovery of microscopy and its application into medical field enabled a more in-depth description of renal vein thrombosis, renal anaemia, acute nephritis, nephrotic syndrome and pyelonephritis for the first time⁷⁶.

However, it has to be taken into account that during the first half of the 20th century the study of impaired human kidneys was reduced to post-mortem autopsy examination of the histology. Moreover, light microscopic findings were often subjected to autolysis, passive ingestion and thick sectioning of the tissue. For that reason, most of kidney affections were advanced and even chronic at the time of the autopsy, so that little to nothing could be elucidated about the early stages and evolution of disease⁷⁷. Afterwards, in the 1940's serum creatinine was advocated as the clinical indicator of renal function by Thomas Addis, using alkaline picrate reaction with plasma samples and matching them by naked eye with a series of standards⁷⁸. This method has evolved a lot due to new technological improvements, though creatinine is the biomarker that eventually prevailed in nephrology practice.

Nowadays, glomerular filtration rate (GFR) is considered the most important measurement of renal function for the assessment of CKD degree⁷⁹. There is not any simple and practical way to measure GFR directly, thus it is an estimated measurement⁸⁰. As kidneys undergo glomerular filtration, tubule reabsorption, tubule secretion and intrarenal metabolism, there is a need for an "ideal GFR marker"⁷⁹. Inulin, iothamylate, radioactively tagged diethylene-triamine penta-acetic acid (⁹⁹Tc-DTPA), radioactively tagged ethylene-diamine tetra-acetic acid (⁵¹Cr-EDTA) or iohexol administrations and urine collections are considered "gold standards" for measuring GFR. However, these exogenous markers are impractical, expensive, time consuming and invasive in the clinical setting, so they are seldom used^{81,82}. In addition, DTPA and EDTA are inefficient in neonates⁸³. Thus, efforts are being made towards the use of endogenous markers in clinical practice.

Serum and/or urine creatinine concentration are currently measured to calculate the clearance of endogenously produced creatinine to be able to estimate GFR⁷⁹. Creatinine, which comes from the catabolism of phosphocreatine in skeletal muscle, is filtered through glomerular capillaries and is secreted by tubular cells^{7,84}.

Table 1.6. Glomerular filtration rate (GFR) and plasma creatinine in healthy children and young adults^{84,85}.

Age (year)	GFR (mean ± SD) (mL/min/ 1.73 m ²)	Age (year)	Plasma creatinine (mg/dL)
		< 2	0.4-0.5
3-4	111.2 ± 18.5	2-8	0.5-0.7
5-6	114.1 ± 18.6		
7-8	111.3 ± 18.3		
9-10	110.0 ± 21.6		
11-12	116.4 ± 18.9	9-18	0.6-0.9
13-15	117.2 ± 16.1		
16-32	112.0 ± 13.0		

The estimation of GFR by means of creatinine clearance involves the need for an accurately timed collection of urine over a long period and is not possible in children who are not toilet trained. For an easier GFR determination, different equations have been created to estimate creatinine clearance from serum creatinine concentration (SCr), including information regarding age, race, gender and anthropometric measurements such as height, weight or body surface area (BSA)^{79,80} (see Table 1.7).

Table 1.7. Different equations to estimate glomerular filtration rate (expressed as mL/min/1.73 m²) and glomerular filtration (mL/min) using serum creatinine^{84,86-88}.

Name	Equation	Remarks
Schwartz	$GFR = k \times L / SCr$	k is an empirical constant: 0.45 for infants in their first year of life, 0.55 for children and adolescent girls and 0.7 for adolescent boys
Counahan-Barratt	$GFR = 0.43 \times L / SCr$	
Modification of Diet in Renal Disease (MDRD)	$GFR = 186 \times SCr^{-1.54} \times age^{-0.203}$ (x 1.21 ^{African-American}) (x 0.742 ^{female})	In case the patient is African-American the result is multiplied by 1.21 and/or if it is a female it is multiplied by 0.742 as well
Cockroft-Gault	$GF = [140 - age] \times weight / [72 \times SCr] \times 0.85^{female}$	If the patient is a female, the result is multiplied by 0.85
*Common remarks:	<ul style="list-style-type: none"> - L is height in cm - Weight is given in kg - SCr is serum creatinine in milligrams per deciliter - The unit for age is years 	

Despite the fact that several equations have been developed and adjusted using large populations, equations based on serum creatinine alone are not accurate enough for GFR estimation. Indeed, serum creatinine is affected by several factors like age, sex, overall body weight and muscular mass, muscular metabolism, nutrition and hydration status⁸⁹, which is the reason why there is a need for including some of these parameters to correct GFR equations. In addition, creatinine lacks sensitivity since serum creatinine levels may keep unaltered until 50 % of the renal function is already lost due to an important nephronic loss^{21,22}. Besides, acute changes in GFR are not represented by SCr until a steady state is achieved and tubular secretion of creatinine may result in overestimation of renal function at low GFR⁹⁰.

For all these reasons, there is a need for new biomarkers to be able to obtain more information about the pathogenesis and evolution of the disease as well as to build new equations useful to diagnose and follow the progression of paediatrics suffering from CKD.

1.9. BIOMARKERS. COMPOUNDS OF INTEREST FOR THE EARLY DETECTION AND INTERVENTION ON THE DISEASE

Biological markers or biomarkers have been defined as cellular, biochemical or molecular alterations that are measurable in biological media such as human tissues, cells or fluids⁹¹. Biomarkers disclosing prodromal or initial signs of a disease enable an earlier diagnosis as well as a better outcome because diagnosis is usually carried out in a more primitive stage of disease⁹². In this sense, creatinine is currently the classic biomarker for CKD both in paediatrics and in adults. However, as previously mentioned in section 1.8, several factors limit the accuracy of creatinine as a GFR marker. For that reason, a vast amount of research is being conducted with the aim of finding new biomarkers for CKD.

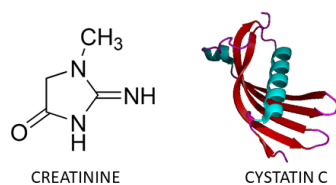


Fig. 1.5. Representation of the structures of creatinine and cystatin C biomarkers of CKD.

Recent investigation in proteomics found that cystatin C is an interesting protein for the diagnosis of CKD, as it is less dependent than creatinine upon patients' body composition, it is not affected by gender and its production rate is usually constant^{93,94}. Accordingly, it

has recently been included in different equations for the calculation of GFR in children aged 1-16 years and GFR ranging from 15 to 75 mL/min/1.73m², together with height, plasma creatinine, blood urea nitrogen (BUN) and gender⁹⁵. However, cystatin C has also some limitations like higher serum levels in elderly, greater weight or height patients, cigarette smoking patients, patients following corticosteroid treatment, hyperthyroidism and malignancy, and lower levels in hypothyroidism⁹³. Therefore, there is a need for finding new compounds to diagnose and follow the progression of CKD.

Another issue to be considered is that, even if biomarkers discovered for adults have commonly been used in clinical practice for both paediatrics and adults, there is little information about the effect of using adult biomarkers in paediatrics⁹⁶. Indeed, human development is known to be a complex process, as ontogeny influences organ function, drug disposition and clinical response to therapy⁹⁷. For example, despite the fact that nephrogenesis is completed by 36 weeks gestational age, during postnatal period renal development continues and several physiological changes occur weeks to months after birth⁹⁸. Hence, creatinine plasma concentration changes with time and follows different patterns in prematures, neonates, children and adults⁹⁹. Thus, it is necessary to address the gap in paediatric biomarkers, taking into account well-recognized developmental changes in children¹⁰⁰.

An ideal paediatric biomarker should comply with the following requirements¹⁰⁰:

1. To be non-invasive
2. To be applicable to specific paediatric diseases
3. To be cost-effective
4. To have well-established paediatric normative values
5. To show a correspondence between results and age-dependent physiologic changes

The use of paediatric CKD-specific biomarkers at an early disease stage would enable prompt diagnosis of specific renal diseases, improving therapeutic treatment and reducing both personal and financial burdens⁸⁹.

1.10. BIBLIOGRAPHY

1. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, Eckardt KU, et al. Chronic kidney disease as a global public health problem: approaches and initiatives - a position statement from Kidney Disease Improving Global Outcomes. *Kidney Int.* 2007; 72: 247-259.
2. Centers for Disease Control and Prevention. Chronic Kidney Disease: Issue Brief 2012.
3. National Kidney Foundation (2002). K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis.* 39: S1-266.
4. Abdomen. In: Drake RL, Vogl W, Mitchell AWM, editors. *Gray Anatomía para estudiantes.* 1st ed: Elsevier; 2005; p. 320-329.
5. Abdomen. In: Moore KL, Dalley AE, editors. *Anatomía con orientación clínica.* 4th ed: Editorial Médica Panamericana; 2002; p. 284-363.
6. Blaus B. Blausen.com staff. Medical gallery of Blausen Medical. *WikiJournal of Medicine* 1(2); 2014.
7. El aparato urinario. In: Tortora GJ, Derrickson B, editors. *Principios de Anatomía y Fisiología.* 11th ed: Editorial Médica Panamericana; 2011; p. 999-1042.
8. Rosner MH. An Overview of Renal Physiology. In: Chapple CR, Steers WD, editors. *Practical Urology: Essential Principles and Practice.* 1st ed, 2011; p. 105-114.
9. Lopez-Novoa JM, Rodriguez-Pena AB, Ortiz A, Martinez-Salgado C, Lopez HFJ. Etiopathology of chronic tubular, glomerular and renovascular nephropathies: clinical implications. *J Transl Med.* 2011; 9: 13.
10. Couser W. Pathogenesis of glomerular damage in glomerulonephritis. *Nephrol Dial Transpl.* 1998; 13: 10-15.
11. Isaka Y, Akagi YA, Y. Cytokines and glomerulosclerosis. *Nephrol Dial Transpl.* 1999; 14: 30-32.
12. United States Renal Data System Coordinating Center. *USRDS Annual Data Report 2016. Epidemiology of kidney disease in the United States.* 2016.
13. Kaspar CDW, Bholah R, Bunchman TE. A Review of Pediatric Chronic Kidney Disease. *Blood Purif.* 2016; 41: 211-217.
14. Bacchetta J, Ranchin B, Demede D, Allard L. The consequences of pediatric renal transplantation on bone metabolism and growth. *Curr Opin Organ Transplant.* 2013; 18: 555-562.
15. Thomas R, Kanso A, Sedor JR. Chronic kidney disease and its complications. *Prim Care.* 2008; 35: 329-344.
16. Becherucci F, Roperto RM, Materassi M, Romagnani P. Chronic kidney disease in children. *Clin Kidney J.* 2016; 9: 583-591.
17. Hogg RJ, Furth S, Lemley KV, Portman R, Schwartz GJ, Coresh J, et al. National Kidney Foundation's Kidney Disease Outcomes Quality Initiative clinical practice

- guidelines for chronic kidney disease in children and adolescents: evaluation, classification, and stratification. *Pediatrics*. 2003; 111: 1416-1421.
18. Levey AS, Eckardt KU, Tsukamoto Y, Levin A, Coresh J, Rossert J, et al. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int*. 2005; 67: 2089-2100.
 19. Agarwal R. Defining end-stage renal disease in clinical trials: a framework for adjudication. *Nephrol Dial Transpl*. 2016; 31: 864-867.
 20. Coresh J, Astor BC, Greene T, Eknoyan G, Levey AS. Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *Am J Kidney Dis*. 2003; 41: 1-12.
 21. Meguid ENA, Bello AK. Chronic kidney disease: the global challenge. *Lancet*. 2005; 365: 331-340.
 22. Hallan SI, Coresh J, Astor BC, Asberg A, Powe NR, Romundstad S, et al. International comparison of the relationship of chronic kidney disease prevalence and ESRD risk. *J Am Soc Nephrol*. 2006; 17: 2275-2284.
 23. European Renal Association - European Dialysis and Transplant Association. ERA-EDTA Registry. 2012.
 24. Unidad de Información sobre pacientes renales de la CAPV, UNIPAR. Servicio Central de Publicaciones del Gobierno Vasco. 2011.
 25. Harambat J, van Stralen KJ, Kim JJ, Tizard EJ. Epidemiology of chronic kidney disease in children. *Pediatr Nephrol*. 2012; 27: 363-373.
 26. Warady BA, Chadha V. Chronic kidney disease in children: the global perspective. *Pediatr Nephrol*. 2007; 22: 1999-2009.
 27. Ardissino G, Dacco V, Testa S, Bonaudo R, Claris-Appiani A, Taioli E, et al. Epidemiology of chronic renal failure in children: data from the Italkid project. *Pediatrics*. 2003; 111: e382-387.
 28. Chesnaye N, Bonthuis M, Schaefer F, Groothoff JW, Verrina E, Heaf JG, et al. Demographics of paediatric renal replacement therapy in Europe: a report of the ESPN/ERA-EDTA registry. *Pediatr Nephrol*. 2014; 29: 2403-2410.
 29. Noordzij M, Kramer A, Abad Diez JM, Alonso de la Torre R, Arcos Fuster E, Bikbov BT, et al. Renal replacement therapy in Europe: a summary of the 2011 ERA-EDTA Registry Annual Report. *Clin Kidney J* 2014; 7: 227-238.
 30. Hughson MD, Douglas-Denton R, Bertram JF, Hoy WE. Hypertension, glomerular number, and birth weight in African Americans and white subjects in the southeastern United States. *Kidney Int*. 2006; 69: 671-678.
 31. McClellan W, Warnock DG, McClure L, Campbell RC, Newsome BB, Howard V, et al. Racial differences in the prevalence of chronic kidney disease among participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Cohort Study. *J Am Soc Nephrol*. 2006; 17: 1710-1715.
 32. Contreras G, Lenz O, Pardo V, Borja E, Cely C, Iqbal K, et al. Outcomes in African Americans and Hispanics with lupus nephritis. *Kidney Int*. 2006; 69: 1846-1851.

33. Eddy A. Pathophysiology of Progressive Renal Disease. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric nephrology*. 6th ed, 2009; p. 1631-1659.
34. Shatat IF, Flynn JT. Hypertension in children with chronic kidney disease. *Adv Chronic Kidney Dis*. 2005; 12: 378-384.
35. Bock JS, Gottlieb SS. Cardiorenal syndrome: new perspectives. *Circulation*. 2010; 121: 2592-2600.
36. Ardissino G, Avolio L, Dacco V, Testa S, Marra G, Vigano S, et al. Long-term outcome of vesicoureteral reflux associated chronic renal failure in children. Data from the ItalKid Project. *J Urol*. 2004; 172: 305-310.
37. Ardissino G, Testa S, Dacco V, Vigano S, Taioli E, Claris-Appiani A, et al. Proteinuria as a predictor of disease progression in children with hypodysplastic nephropathy. Data from the Ital Kid Project. *Pediatr Nephrol*. 2004; 19: 172-177.
38. Serra A, Romero R, Lopez D, Navarro M, Esteve A, Perez N, et al. Renal injury in the extremely obese patients with normal renal function. *Kidney Int*. 2008; 73: 947-955.
39. Hsu CY, McCulloch CE, Iribarren C, Darbinian J, Go AS. Body mass index and risk for end-stage renal disease. *Ann Intern Med*. 2006; 144: 21-28.
40. Schwarz S, Trivedi BK, Kalantar-Zadeh K, Kovesdy CP. Association of disorders in mineral metabolism with progression of chronic kidney disease. *Clin J Am Soc Nephrol*. 2006; 1: 825-831.
41. Muntner P, He J, Astor BC, Folsom AR, Coresh J. Traditional and nontraditional risk factors predict coronary heart disease in chronic kidney disease: results from the atherosclerosis risk in communities study. *J Am Soc Nephrol*. 2005; 16: 529-538.
42. Saland JM, Pierce CB, Mitsnefes MM, Flynn JT, Goebel J, Kupferman JC, et al. Dyslipidemia in children with chronic kidney disease. *Kidney Int*. 2010; 78: 1154-1163.
43. North American Pediatric Renal Transplant Cooperative Study (NAPRTCS) 2007 Annual Report. The EMMES Corporation.
44. Castaneda DA, Lopez LF, Ovalle DF, Buitrago J, Rodriguez D, Lozano E. Growth, Chronic Kidney Disease and Pediatric Kidney Transplantation: Is It Useful to Use Recombinant Growth Hormone in Colombian Children With Renal Transplant? *Transplant Proc*. 2011; 43: 3344-3349.
45. Sahay M, Sahay R, Kalra S, Baruah MP. Nutrition in chronic kidney disease. *J Med Nutr Nutraceuticals*. 2014; 3: 11-18, 18 pp.
46. Gonzalez EA, Sachdeva A, Oliver DA, Martin KJ. Vitamin D insufficiency and deficiency in chronic kidney disease. A single center observational study. *Am J Nephrol*. 2004; 24: 503-510.
47. LaClair RE, Hellman RN, Karp SL, Kraus M, Ofner S, Li Q, et al. Prevalence of calcidiol deficiency in CKD: a cross-sectional study across latitudes in the United States. *Am J Kidney Dis*. 2005; 45: 1026-1033.

48. Jofre R, Rodriguez-Benitez P, Lopez-Gomez JM, Perez-Garcia R. Inflammatory syndrome in patients on hemodialysis. *J Am Soc Nephrol*. 2006; 17: S274-S280.
49. Menon V, Greene T, Wang X, Pereira AA, Marcovina SM, Beck GJ, et al. C-reactive protein and albumin as predictors of all-cause and cardiovascular mortality in chronic kidney disease. *Kidney Int*. 2005; 68: 766-772.
50. Matevossian E, Kern H, Huser N, Doll D, Snopok Y, Nahrig J, et al. Surgeon Yurii Voronoy (1895-1961) - a pioneer in the history of clinical transplantation: in memoriam at the 75th anniversary of the first human kidney transplantation. *Transpl Int*. 2009; 22: 1132-1139.
51. Blagg CR. The early history of dialysis for chronic renal failure in the United States: a view from Seattle. *Am J Kidney Dis*. 2007; 49: 482-496.
52. Drukker W. History of peritoneal dialysis. In: Maher JF, editor. *Replacement of Renal Function by Dialysis*. 3rd ed. Dordrecht, The Netherlands: Kluwer; 1987; p. 475-515.
53. Tenckhoff H, Shilipetar G, Van PWH, Swanson E. A home peritoneal dialysate delivery system. *Trans Am Soc Artif Intern Organs*. 1969; 15: 103-107.
54. Perico N, Benigni A, Remuzzi G. Present and future drug treatments for chronic kidney diseases: evolving targets in renoprotection. *Nat Rev Drug Discovery*. 2008; 7: 936-953.
55. Peterson JC, Adler S, Burkart JM, Greene T, Hebert LA, Hunsicker LG, et al. Blood pressure control, proteinuria, and the progression of renal disease. The Modification of Diet in Renal Disease Study. *Ann Intern Med*. 1995; 123: 754-762.
56. Wright JT, Jr., Bakris G, Greene T, Agodoa LY, Appel LJ, Charleston J, et al. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease. Results from the AASK trial. *JAMA, J Am Med Assoc*. 2002; 288: 2421-2431.
57. Jafar TH, Schmid CH, Landa M, Giatras I, Toto R, Remuzzi G, et al. Angiotensin-converting enzyme inhibitors and progression of nondiabetic renal disease: A meta-analysis of patient-level data. *Ann Intern Med*. 2001; 135: 73-87.
58. Van de Voorde RG, Warady BA. Management of Chronic Kidney Disease in Pediatric Nephrology. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 1661-1716.
59. Sica DA. Angiotensin receptor blockers: new considerations in their mechanism of action. *J Clin Hypertens*. 2006; 8: 381-385.
60. Van Dyck M, Proesmans W. Renoprotection by ACE inhibitors after severe hemolytic uremic syndrome. *Pediatr Nephrol*. 2004; 19: 688-690.
61. Tanaka H, Suzuki K, Nakahata T, Tsugawa K, Konno Y, Tsuruga K, et al. Combined therapy of enalapril and losartan attenuates histologic progression in immunoglobulin A nephropathy. *Pediatr Int*. 2004; 46: 576-579.
62. Ardissino G, Vigano S, Testa S, Dacco V, Paglialonga F, Leoni A, et al. No clear evidence of ACEi efficacy on the progression of chronic kidney disease in children

- with hypodysplastic nephropathy--report from the Italkid Project database. *Nephrol Dial Transplant*. 2007; 22: 2525-2530.
63. Epstein M. Evolving therapeutic strategies for retarding progression of diabetic nephropathy--an update for 2002. *Acta Diabetol*. 2002; 39 Suppl 2: S41-45.
 64. Epstein M. Recent landmark clinical trials: how do they modify the therapeutic paradigm? *Am J Hypertens*. 2002; 15: 82S-84S.
 65. Epstein M, Campese VM. Pleiotropic effects of 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors on renal function. *Am J Kidney Dis*. 2005; 45: 2-14.
 66. Sandhu S, Wiebe N, Fried LF, Tonelli M. Statins for improving renal outcomes: a meta-analysis. *J Am Soc Nephrol*. 2006; 17: 2006-2016.
 67. Atkinson MA, Furth SL. Anemia in children with chronic kidney disease. *Nat Rev Nephrol*. 2011; 7: 635-641.
 68. Wingen AM, Fabian-Bach C, Mehls O. Low-protein diet in children with chronic renal failure--1-year results. European Study Group for Nutritional Treatment of Chronic Renal Failure in Childhood. *Pediatr Nephrol*. 1991; 5: 496-500.
 69. Graf L, Nailescu C, Kaskel PJ, Kaskel FJ. Nutrition and Metabolism. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 307-323.
 70. Fleming GM. Renal replacement therapy review: past, present and future. *Organogenesis*. 2011; 7: 2-12.
 71. Verrina E. Peritoneal Dialysis. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 1785-1816.
 72. Slinin Y, Greer N, MacDonald R, Rutks I, Ishani A, Wilt TJ, et al. Timing of dialysis initiation, duration and frequency of hemodialysis sessions, and membrane flux: a systematic review for a KDOQI clinical practice guideline. *Am J Kidney Dis*. 2015; 66: 823-836.
 73. Ingulli E, Alexander SI, Briscoe DM. Transplantation Immunobiology. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 1835-1866.
 74. Marketos SG. Hippocratic medicine and nephrology. *Am J Nephrol*. 1994; 14: 264-269.
 75. Weening JJ, Jennette JC. Historical milestones in renal pathology. *Virchows Arch*. 2012; 461: 3-11.
 76. Richet G. From Bright's disease to modern nephrology: Pierre Rayer's innovative method of clinical investigation. *Kidney Int*. 1991; 39: 787-792.
 77. D'Agati VD, Mengel M. The Rise of Renal Pathology in Nephrology: Structure Illuminates Function. *Am J Kidney Dis*. 2013; 61: 1016-1025.
 78. Peitzman SJ. Thomas Addis (1881-1949): mixing patients, rats, and politics. *Kidney Int*. 1990; 37: 833-840.

79. Hunley TE, Kon V, Ichikawa I. Glomerular Circulation and Function. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 46-50.
80. National Kidney Foundation. Cystatin C: what is its role in estimating GFR? 2009.
81. Wasung ME, Chawla LS, Madero M. Biomarkers of renal function, which and when? *Clin Chim Acta*. 2015; 438: 350-357.
82. Swedish Council on Health Technology Assessment. Methods to Estimate and Measure Renal Function: a systematic review. 2013.
83. Friedman A. Laboratory Assessment and Investigation of Renal Function. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 491-504.
84. Schwartz GJ, Work DF. Measurement and estimation of GFR in children and adolescents. *Clin J Am Soc Nephrol*. 2009; 4: 1832-1843.
85. García-Nieto V, Santos F. Pruebas funcionales Renales. In: García-Nieto V, Santos F, editors. *Nefrología Pediátrica*. Madrid: Aula Médica; 2000; p. 15-26.
86. Counahan R, Chantler C, Ghazali S, Kirkwood B, Rose F, Barratt TM. Estimation of glomerular filtration rate from plasma creatinine concentration in children. *Arch Dis Child*. 1976; 51: 875-878.
87. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med*. 1999; 130: 461-470.
88. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976; 16: 31-41.
89. Zhao YY. Metabolomics in chronic kidney disease. *Clin Chim Acta*. 2013; 422: 59-69.
90. De Geus HRH. Biomarkers for AKI: Erasmus MC University Medical Center Rotterdam; 2013.
91. Hulka BS. Overview of biological markers. In: Hulka B, Griffith J, Wilcosky T, editors. *Biological markers in epidemiology*. 1st ed. New York: Oxford University Press; 1990; p. 3-15.
92. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx*. 2004; 1: 182-188.
93. Chew JSC, Saleem M, Florkowski CM, George PM. Cystatin C--a paradigm of evidence based laboratory medicine. *Clin Biochem Rev*. 2008; 29: 47-62.
94. Grubb A. Cystatin C as a Biomarker in Kidney Disease. In: Edelstein CL, editor. *Biomarkers in Kidney Disease*. 1st ed: Academic Press; 2010; p. 291-312.
95. Schwartz GJ, Munoz A, Schneider MF, Mak RH, Kaskel F, Warady BA, et al. New equations to estimate GFR in children with CKD. *J Am Soc Nephrol*. 2009; 20: 629-637.

96. Lanza IR, Zhang S, Ward LE, Karakelides H, Raftery D, Nair KS. Quantitative metabolomics by H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. *PLoS One*. 2010; 5: 1-10.
97. Kearns GL, Abdel-Rahman SM, Alander SW, Blowey DL, Leeder JS, Kauffman RE. Developmental pharmacology-drug disposition, action, and therapy in infants and children. *N Engl J Med*. 2003; 349: 1157-1167.
98. Chevalier RL. Developmental renal physiology of the low birth weight pre-term newborn. *J Urol*. 1996; 156: 714-719.
99. Shin S, Fung S-M, Mohan S, Fung H-L. Simultaneous bioanalysis of L-arginine, L-citrulline, and dimethylarginines by LC-MS/MS. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2011; 879: 467-474.
100. Goldman J, Becker ML, Jones B, Clements M, Leeder JS. Development of biomarkers to optimize pediatric patient management: what makes children different? *Biomarkers Med*. 2011; 5: 781-794.



CHAPTER II.

OBJECTIVES

The prospects for the clinical management of CKD should be analyzed from different perspectives. First, diagnosis improvement is necessary by means of an earlier detection of the disease as well as by the identification of progression and prognostic markers to be able to determine which patients will progress faster than others will. In addition, new therapies or treatments are needed to be able to improve the existing capacity to slow or even stop the progression of CKD and both reverse the damage and regenerate impaired renal tissues.

As mentioned in the first chapter of this work, there is a need for new biomarkers capable of complying with an earlier diagnosis of CKD and following the progression of the disease. Moreover, this requirement is even more necessary in paediatrics, as the etiopathology of the disease differs between adults and paediatrics and both the lack of information and research in paediatric population are even more noteworthy.

In order to achieve this goal, the present study is clearly structured around three operational objectives:

1. Targeted metabolomic study focused on amino acids, amino acid derivatives and related compounds from arginine-creatinine, arginine methylation and urea cycle suspicious of being altered in paediatrics with CKD.
2. Untargeted metabolomic study for unbiased biomarker discovery.
3. A study based on the significant biomarkers found in the previous two sections to validate the potential biomarkers from the untargeted research as well as to be able to quantify all of these biomarkers together.

Plasma samples from paediatric CKD patients and controls from the Basque Country, collected by Paediatric Nephrology Service at Cruces Hospital (Barakaldo, Spain) will be used to comply with all these scopes.

Concerning the first operational objective, due to the nature of some amino acids, it deals with a specific sample treatment method development including the use of reducing agents to be able to quantify them. In addition, amino acids and amino acid derivatives are small and polar metabolites. Therefore, an ion-pairing reversed-phase liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) method must be established in plasma samples from both control and CKD paediatrics for 16 metabolites.

As far as the second operational objective is concerned, the extraction of plasma samples by means of protein precipitation with different solvent combinations is required in order to obtain as many entities as possible without any bias. In addition, a LC-QTOF-MS method has to be developed to perform the untargeted metabolomics approach. Then, in-depth univariate and multivariate data analysis will be highly necessary to deal with the complex data set expected from the untargeted metabolomics analysis in order to obtain the significant features, which may enable differentiating between CKD and control patients.

Finally, the last operational objective deals with the development of both specific plasma extraction procedure and liquid chromatography coupled to triple quadrupole mass spectrometry (LC-QQQ-MS) method to quantify the significant analytes obtained from the first and second scopes. The aim of this method is the validation of the results from the untargeted metabolomics method and, at the same time, the quantification of these metabolites together to get a new analytical method that may result in better diagnosis and monitoring of CKD patients in comparison with traditional methods.



CHAPTER III.

METHODOLOGY

- 3.1. Introduction**
- 3.2. Metabolomics**
- 3.3. Analytical instrumentation**
 - 3.3.1. Nuclear magnetic resonance (NMR)**
 - 3.3.2. Mass spectrometry (MS)**
- 3.4. Statistical and chemometrical approach**
 - 3.4.1. Hierarchical clustering**
 - 3.4.2. Principal component analysis (PCA)**
 - 3.4.3. Classification methods**
 - 3.4.4. Sparse partial least squares discriminant analysis (SPLS-DA)**
- 3.5. Bibliography**

3.1. INTRODUCTION

The advent of high-throughput technologies has revolutionized biomedical research by means of “omics” technologies, including genomics, transcriptomics, proteomics and metabolomics, which are speeding up the discovery of new biomarkers for different diseases².

Over the years both biomedical research and the progress of technology have been directly correlated with improvements in comprehension, diagnosis, follow-up and treatment of chronic kidney disease (CKD). Moreover, life quality and expectancy of patients suffering from CKD have improved due to all these advances. This chapter intends to explain the methodological and technological novelties that have been considered in this work with the aim of finding new biomarkers for paediatric CKD diagnosis and monitoring in plasma.

3.2. METABOLOMICS

“Omic” sciences are primarily aimed at the universal detection and characterization of genes (genomics), *m*RNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample and at the adoption of a holistic view of the molecules that make up a cell, tissue or organism³.

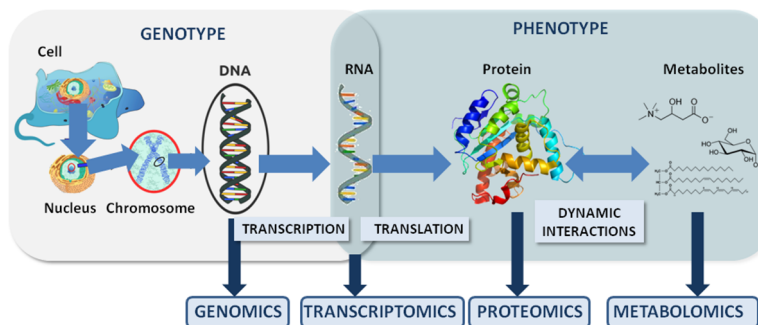


Fig. 3.1. Schematic representation of “omic” sciences.

Metabolomics is the systematic study of small molecules (typically below 1500 Da) or metabolites, in a biological system. Metabolites, defined as the end products of cellular regulatory processes, are considered the latest response of biological systems to genetic or environmental changes⁴. These compounds include amino acids and biogenic amines, small peptides, carbohydrates, lipids, nucleic acids, vitamins, organic acids and intermediary energy metabolites, for instance⁵.

The feasibility of metabolomics for biomarker discovery is sustained by the premise that metabolites have an important role in biological systems and that diseases originate from disruption of metabolic pathways⁶. It is considered that metabolomics is more advantageous over other "omic" approaches, as metabolites are the latest downstream outputs of gene transcription. Accordingly, metabolites act as direct signatures of biochemical activity, changes in the metabolome are amplified compared with the transcriptome and the proteome and therefore, the correlation with phenotype is more straightforward^{3,7}. In addition, the number of metabolites is lower than the number of genes and proteins in a cell, so sample complexity is reduced. Although the concentration of enzymes and metabolic flux may not significantly change during a biochemical reaction, the concentration of the metabolites can significantly fluctuate⁸. A great variety of metabolites, such as blood cholesterol, glucose or triglycerides have been successfully applied as measured biomarkers of different pathologies for decades⁹.

Metabonomics is a related term to define changes in metabolic activities occurred as a response of living systems to pathophysiological stimuli or genetic modifications¹⁰. There are also some other useful working definitions, such as, metabolic profiling, metabolic fingerprinting or metabolic footprinting. The first one refers to the identification and quantification of a number of metabolites generally related to specific metabolic pathways or to specific classes of compounds^{8,11}. Metabolic fingerprinting is a high-throughput, rapid, global analysis of samples to provide sample classification, with no need for either quantification or metabolite identification^{8,12}. On the contrary, metabolic footprinting refers to the measurement of metabolites secreted by a cell or a tissue, rather than intracellular metabolites^{13,14}.

Some other subdisciplines have emerged due to the complexity of the metabolism, such as, lipidomics or nutrimental metabolomics. Lipidomics, refers to the study of the lipid fraction of the metabolome using metabolomic approaches⁹. Nutrimental metabolomics aims to identify biomarkers that allow the monitorization of dietary components that play a role in health maintenance¹⁵.

According to the strategy followed, metabolomic analyses are generally classified into two complementary approaches: targeted or untargeted metabolomics. Both approaches can provide new clinically useful indicators or biomarkers to diagnose health status, disease or the possible outcomes of a defined treatment.

On the one hand, targeted metabolomics refers to an strategy in which a previously defined list of metabolites is measured motivated by a specific biochemical question or hypothesis^{7,16}. These pre-defined metabolites can be associated with specific pathways, enzyme systems or a complete metabolite class⁵. Particularities of targeted metabolomics are explained in detail in Chapter 4.

On the other hand, untargeted metabolomics consists of an unbiased holistic approach towards simultaneously analyzing as many features as possible after minimal sample pretreatment, without any prior knowledge of the identity of these features and any bias^{5,7,17}. Chapter 5 provides more information about the characteristics of untargeted metabolomics studies.

3.3. ANALYTICAL INSTRUMENTATION

Metabolomics has become possible through the arrival of recent technological breakthroughs in small-molecule separation and identification¹⁸. Nuclear magnetic resonance (NMR) is a well-established powerful analytical technique for metabolomics¹⁹. However, the trend is towards the use of mass spectrometry (MS).

3.3.1. NUCLEAR MAGNETIC RESONANCE (NMR)

NMR is a reproducible and fast spectroscopic technique based on the absorption of energy and re-emission of the atom nuclei due to variations in an external magnetic field²⁰. ¹H (proton), ¹³C, ¹⁵N and ³¹P are the most widely used nuclei in biomolecular NMR, being one-dimensional (1D) ¹H NMR the most frequently used in metabolomics due to its natural abundance in biological samples^{21,22}. Applications of NMR are exponentially growing with the refinement of measurement methods, analysis and interpretation of complex data sets²³. NMR is appropriate for molecules with poor ionization properties, such as glucose¹⁷. Sample preparation is less time-consuming in NMR in comparison with gas chromatography coupled to mass spectrometry (GC-MS) and liquid chromatography coupled to mass spectrometry (LC-MS). However, GC-MS and LC-MS achieve higher sensitivity than NMR, so that metabolites present in a concentration below the detection limit of NMR may be detected using these techniques^{19,24}.

3.3.2. MASS SPECTROMETRY (MS)

Mass spectrometry (MS) is a well-established detection method, which allows the simultaneous detection of multiple analytes with high sensitivity. This instrumental technique has raised to an outstanding position among analytical methods, also in the field of metabolomics, due to its unequalled sensitivity allowing unequivocal identification of a substance, detection limits, speed, universality and diversity of its applications^{25,26}.

Mass spectrometry is based on the following assumptions: atoms or molecules from a sample are ionized, separated according to its mass-to-charge relation (m/z) through a magnetic or electrical field and are subsequently detected and registered in mass spectra. The mass spectrum is a graphical display representing the relative abundance of ion signals against the m/z ratios²⁷. To comply with its function, mass spectrometers consist of an ionization source, a mass analyzer and a detector.

From an analytical point of view, metabolomics involves a challenge, due to the complexity of biological samples. For that reason, both in targeted and untargeted metabolomics, when using mass spectrometry for detection, this is generally preceded by a separation step, which reduces the high complexity of biological samples, allows MS analysis of different sets of molecules at different times and increases the quality of the raw data obtained^{22,25}. Mass spectrometry is usually coupled to gas chromatography (GC), liquid chromatography (LC) or capillary electrophoresis (CE) techniques. LC-MS is the most commonly used separation analytical technique, because of its sensitivity, adaptability for exact quantitative determinations, suitability for the automatization of the processes, capability to separate nonvolatile or thermolabile species, and especially, its applicability to a wide variety of substances important for the industry, science and society²⁸. Due to its high throughput and the good coverage of metabolites in different bio fluids, the popularity of LC-MS platform for metabolomic studies is growing²⁹. GC-MS and CE-MS have been considered complementary to LC-MS in terms of increasing the metabolome coverage and to verify the identification of potential biomarkers found by LC-MS^{30,31}.

The main obstacle for LC-MS-based techniques has been the incompatibility of the liquid eluent coming from the column and the vacuum of the mass spectrometer. Indeed, one of the prerequisites for mass spectrometric analysis is the formation of volatilized ions, so that there is a need for elimination of the mobile phase. With the aim of solving these problems, several ion-source coupling methods have been developed³². Regarding mass analyzer, it deals with ion separation according to their mass-to-charge ratio (m/z).

Finally, the detector is the source responsible for transforming each ion m/z relation into an electrical signal (see Fig. 3.2).

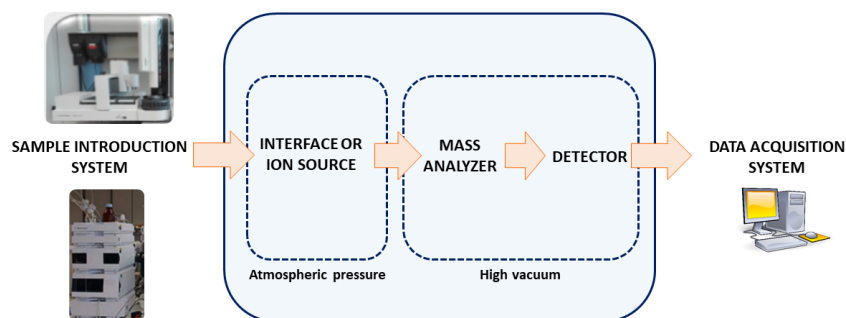


Fig. 3.2. Schematic diagram of a mass spectrometer.

The following sections will focus especially on the most relevant aspects of LC-MS, selected for the development of the different analytical methods in this work.

3.3.2.1. Ion sources

Molecules coming from liquid chromatography system must be in a gaseous form and ionized to be analyzed in a mass spectrometer. Indeed, ion production by different ion sources often implies gas-phase ion-molecule reactions²⁶.

Vacuum inside the instrument must remain unchanged when sample is introduced into the ionization source. Depending on the sample type and the ionization source, samples can be introduced by direct infusion or by direct insertion into the mass spectrometer, and when these sources work at atmospheric pressure, they are known as atmospheric pressure ionization sources (API)²⁶. Theoretically, the selection of the ionization source relies on the chemical properties of the targeted analyte. However, when untargeted metabolomics experiments are carried out, as the metabolites of interest are unknown, the use of different interfaces is recommended. Figure 3.3 schematizes the applicability of LC-MS ionization techniques according to the polarity and molecular weight of the analytes of interest.

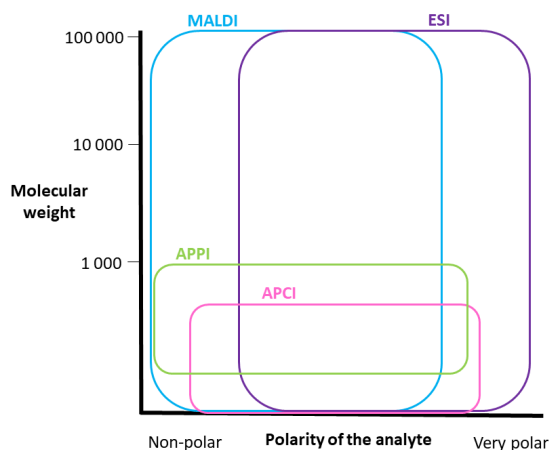


Fig. 3.3. Applicability of different ionization sources in LC-MS.

The most commonly used ion sources are detailed in the following subsections:

✓ **Electrospray interface (ESI)**

Electrospray ionization or ESI is the most universally known ionization method with very low chemical specificity and therefore, it is the most widely used ionization technique in biochemical field for the analysis of samples in liquid form³³. Moreover, ESI coupled to liquid chromatography and together with mass spectrometry has become a very powerful technique even in complex biological samples²⁷.

ESI works by applying a strong electric field to a previously ionized liquid passing through a capillary tube with a low flux (1-10 $\mu\text{L}/\text{min}$). This electric field is achieved by handling a potential difference between the capillary and the skimmer, and as a result, highly charged droplets with the same polarity as the capillary voltage are generated at the end of the capillary^{27,33}. In addition, a gas is injected coaxially at a low flow rate and afterwards, these droplets pass either through a curtain of heated inert gas (usually nitrogen) or through a heated capillary to remove the last solvent molecules³⁴. Charged droplets are continuously reduced in size due to the evaporation of the solvent, so that there is an increase of surface charge density and a decrease of the droplet radius. Finally, electrostatic repulsion becomes stronger than the surface tension and when Rayleigh instability limit is exceeded a small electrically charged droplet leaves the surface and is ejected into the gaseous phase²⁷.

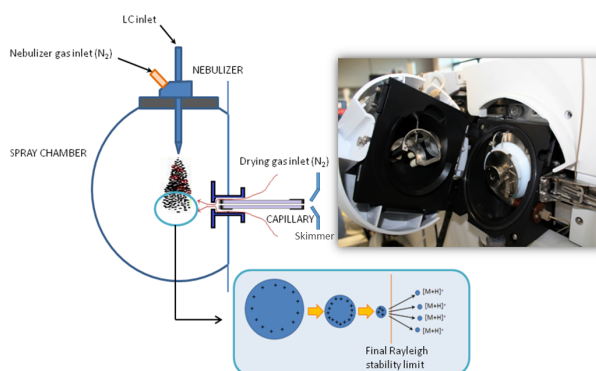


Fig. 3.4. Electro spray ionization source (ESI).

Analyses can be performed in ESI positive and ESI negative modes, thus obtaining characteristic $[M+H]^+$ and $[M-H]^-$ ions, respectively. The adjustment of pH of the mobile phase is essential when ESI is interfaced to LC, as the liquid needs to be previously ionized. Positive and negative ion modes require different solvent characteristics and the response of a given analyte can be enhanced or suppressed using different solvent systems. Volatile weak acid and base buffers are preferred, being the most common choices formic acid, ammonium acetate and ammonium formate³⁵.

ESI is a soft ionization technique in the sense that a very little residual energy is retained by the analyte and generally no fragmentation occurs upon ionization^{32,36}. In addition, it allows very high sensitivity, it is robust and a reliable tool for studying non-volatile and thermally labile biomolecules that cannot be analyzed by other conventional techniques and is easy to couple to LC³⁴.

✓ **Atmospheric Pressure Chemical Ionization interface (APCI)**

Atmospheric Pressure Chemical Ionization or APCI is an ionization technique based on the use of gas-phase ion-molecule reactions at atmospheric pressure and is mainly used for polar and relatively nonpolar compounds with moderate molecular weight up to 1500 Da, and giving generally monocharged ions³⁴.

The performance of APCI is optimal at high flow-rates (1 mL/min and higher)³². The solution coming from the column is nebulized by a heated nitrogen flow coaxial to the capillary and the solvent is evaporated rapidly with the help of the heated vaporizer tube. In addition, the gas-vapour mixture contacts a corona discharge

needle, which is held at a voltage of 2.5 to 3 kV, causing around the needle plasma containing charged particles and free electrons^{37,38}. APCI is intended for analyzing weakly polar or neutral analytes³⁹.

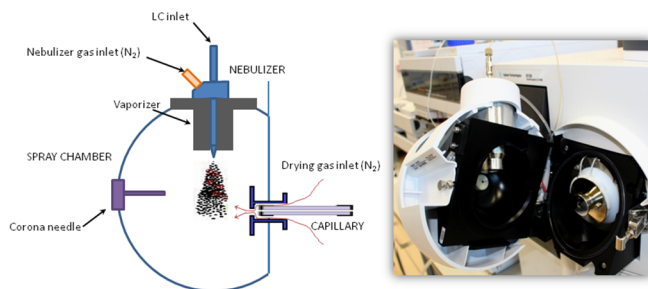


Fig. 3.5. Atmospheric Pressure Chemical Ionization source (APCI).

✓ Atmospheric pressure photoionization (APPI)

Atmospheric pressure photoionization or APPI is becoming popular as it is able to ionize compounds that are often missed or not sufficiently ionized by ESI and APCI⁴⁰. APPI requires less heat for desolvation than APCI, thus, it is useful for the analysis of thermally labile compounds, increasing the range of compounds susceptible to ionization⁴¹.

APPI is based on the use of photons from a light source, typically a krypton lamp, to ionize analyte molecules. Analyte molecules might be ionized by two mechanisms with APPI: direct photoionization and photo-induced chemical ionization. A dopant or a photoionizable reagent (usually acetone or toluene) is added to the mobile phase to favour the ionization by transferring a charge to the analyte⁴¹.

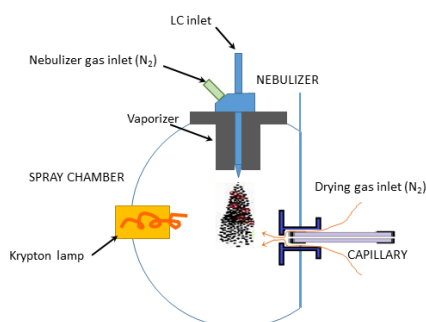


Fig. 3.6. Atmospheric Pressure Photoionization Ionization source (APPI).

✓ **Matrix-Assisted Laser Desorption Ionization interface (MALDI)**

Matrix-Assisted Laser Desorption Ionization or MALDI is a powerful source for the production of intact gas-phase ions for mass spectrometry analysis from large, nonvolatile and thermally labile compounds such as proteins, oligonucleotides, synthetic polymers and large inorganic compounds³⁴.

MALDI involves the use of a suitable solvent where the analyte and the matrix are dissolved, the probe surface and a laser, for which laser wavelength and fluence (energy per area) are adjusted⁴². Laser irradiation heats the crystals rapidly by accumulating a large amount of energy through excitation of the matrix molecules and causes localized sublimation of matrix crystals, ablation of a portion of the crystal surface and the expansion of the matrix into the gas phase, catching intact analyte in the expanding matrix plume⁴³.

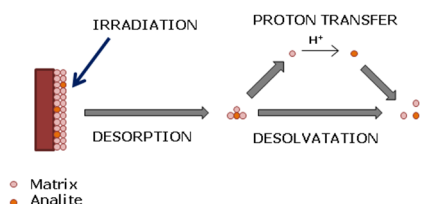


Fig. 3.7. Diagram of the principles of MALDI.

✓ **Multimode interface**

About 80% of all analytes are estimated to respond by ESI ionization and APCI is one of the best-known alternatives to detect analytes that do not ionize by ESI. As a consequence, novel ion sources have been developed combining these two ionization techniques into one source, known as “dual-mode” or “multimode” sources⁴⁴.

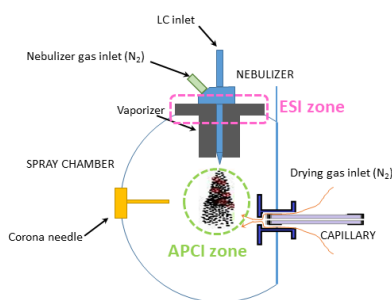


Fig. 3.8. Multimode ion source.

3.3.2.2. Mass analyzers

Mass analyzers work under high vacuum conditions to separate gas-phase ions produced according to their mass-to-charge ratio (m/z). This separation is based on different principles, such as static or dynamic electric and magnetic fields alone or combined. The main characteristics for evaluating the performance of a mass analyzer are the following: resolution, mass accuracy, mass range, analysis speed and lineal dynamic range⁴⁵.

Resolving power (RP) is an essential parameter to characterize the ability of a particular mass analyzer to resolve peaks in mass spectra, defined as the m/z value of a defined peak divided by the peak full width at half maximum (FWHM), as showed in Equation 3.1⁴⁶. FWHM is defined as the peak width of a certain m/z measured at 50 % peak height⁴⁷.

$$RP = \frac{m/z}{\Delta m/z}$$

Eq. 3.1. Resolving power for a particular m/z value.

It has to be noted that RP must be determined for a particular m/z value, as RP grows when increasing m/z value under identical peak width condition.

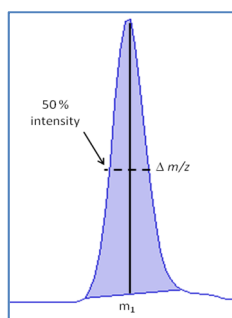


Fig. 3.9. Definition of resolving power according to FWHM.

Mass accuracy or mass measurement error is calculated by subtracting exact mass (calculated mass) to the accurate mass (experimental mass), dividing this result with the exact mass and multiplying it with a million to express it as parts per million⁴⁸, as showed in Equation 3.2.

$$\text{Mass accuracy (ppm)} = \frac{m_{\text{accur}} - m_{\text{exact}}}{m_{\text{exact}}} \times 10^6$$

Eq. 3.2. Mass accuracy expressed as parts per million.

The acquisition speed is the time needed to acquire one spectrum or one data point in a chromatogram and is typically expressed in Da/s for low-resolution and in Hz for high-resolution mass analyzers^{46,49}. Even though mass accuracy is often wanted, it is usually achieved at the expense of lower sensitivity or slower data acquisition speed per ion⁵⁰.

Regarding linear dynamic range, the range over which the relation between ion signal and analyte concentration is linear, it depends not only on the particular instrument but also in the application⁴⁶.

The election of the mass analyzer is fundamentally defined by the working parameters, as showed in Table 3.1.

Table 3.1. Common parameters quadrupole (Q), ion trap (IT), time-of-flight (TOF), Orbitrap and ion cyclotron resonance (ICR) mass analyzers used in LC-MS⁴⁶.

Analyzer	Resolving power (x10 ³)	Mass accuracy (ppm)	m/z range (upper limit) (x10 ³)	Acquisition speed (Hz)	Lineal dynamic range
Q	3-5	Low	2-3	2-10	10 ⁵ -10 ⁶
IT	4-20	Low	4-6	2-10	10 ⁴ -10 ⁵
TOF	10-60	1-5	10-20	10-50	10 ⁴ -10 ⁵
Orbitrap	100-240	1-3	4	1-5	5 x 10 ³
ICR	750-2500	0.3-1	4-10	0.5-2	10 ⁴

The following mass analyzers have been used during the course of this work:

✓ **Quadrupole Analyzer (Q)**

Quadrupole analyzer is the simplest device, which uses the stability of the trajectories in oscillating electric fields to separate ions according to their *m/z* ratios. Quadrupoles are low-resolution instruments, as they are operated at unit resolution, that is, a resolution that is sufficient to separate two peaks one mass unit apart⁴⁵. For that purpose, four perfectly parallel rods of circular or hyperbolic section are used⁴⁵. These rods are placed few centimetres apart and if radio frequency (RF) is applied to adjacent rods, an ion injected along the axis will start spinning in an imaginary cylinder, and depending on the voltage and frequency, the ion will pass through the quadrupole. If direct current voltage (DC) is applied on top of the RF voltage, mass

separation is obtained due to the narrowing down of the transmitted m/z range⁵¹. The diameter of the imaginary cylinder of each ion depends on the mass-to-charge ratio and only ions within a certain m/z range will keep all the way through the quadrupole^{45,51}.

The quadrupole (or hexapole or octapole, depending on the number of rods) may work in Scan mode as a wide pass filter when only RF voltage is applied to the quadrupole, due to a high transmission of ions within a wide m/z range. In addition, quadrupole is used for focusing selected ion beams in Selected Ion Monitoring (SIM) mode, by only filtering ions with defined m/z relations, providing a high sensitivity. Furthermore, quadrupoles are frequently used in combination with other mass analyzers in tandem mass spectrometry (MS/MS) to focus the ion beams as well as in collision cells⁵¹.

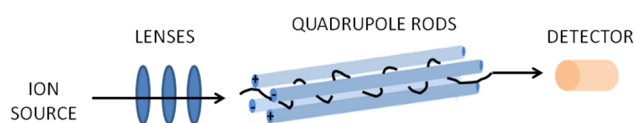


Fig. 3.10. Schematic diagram of a quadrupole mass spectrometer.

✓ Time-of-flight Analyzer (TOF)

Time-of-flight (TOF) analyzers work on the basis of determining ion's mass-to-charge ratio via a time measurement. TOF was first described by Stephens in 1946 during the Spring Meeting of the American Physical Society⁵² and a commercial TOF based on Wiley and McLaren design⁵³ was launched onto the market for the first time in 1956 by Bendix Corporation (USA). Afterwards, in the late 80s and early 90s the groups of Guilhaus and Dodonov introduced the concept of the "orthogonal acceleration" TOF concept, thus improving resolving power, repeatability, linearity and sensitivity⁵⁴.

The operation of TOF is based on the acceleration of ions using an electric field of known strength. As a result, all the ions have the same kinetic energy and the velocity of the ion will depend on the m/z ratio. Therefore, the time needed by an ion to reach the detector at a known distance is measured, which is directly proportional to m/z ratio⁵⁵. As velocity (v) is dependent upon the kinetic energy (E) and mass (m), lighter ions will travel faster and will arrive earlier to the detector⁵⁵.

$$E = \frac{1}{2} m v^2$$

Eq. 3.3.

From Equation 3.3 the mass of a molecule can be calculated by measuring the time a molecule requires to fly a known distance. The velocity of the ion leaving the source is given by rearranging the same equation to give the velocity of an ion (v) as a function of acceleration voltage (V) and m/z value (see Eq. 3.4).

$$v = \sqrt{\left(\frac{2V \times e}{m/z}\right)}$$

Eq. 3.4.

It has to be taken into account that the time delay (t) from ion formation to the time the detector is reached depends on the length of the drift region (L), m/z ratio of the ion and the acceleration voltage in the source.

$$t = \frac{L}{V} = L / \sqrt{\left(\frac{2V \times e}{m/z}\right)}$$

Eq. 3.5.

In practice, this process starts with a so-called *pusher or ion pulser*, where ions enter and a short electric pulse is used to accelerate them to the same kinetic energy and at the same time to start a timer. Then, the ions drift through a flying tube. This flying tube contains an electric mirror or *reflectron*, which reverses the ion beam elongating the flying path and correcting the residual differences in kinetic energy. Finally, after going through the flying tube, which is usually 1 to 2 meter long, the ions reach the detector. Due to the length of the flying tube, a significant increase in mass resolution and mass accuracy is achieved.

In general, TOF analyzers are very fast and up to 20000 push events can be done per second, even if spectra from many push events are then summarized into one mass spectrum to improve ion statistics and reduce noise^{51,55}. In addition to this high acquisition speed, it provides accurate mass measurement with an adequate calibration range as well as full scan spectral sensitivity⁵⁶. Another strength of TOF instruments is their high resolution. Indeed, they generally have 10000 or more resolving power, expressed in terms of full peak width at one-half maximum (FWHM)⁴⁷.

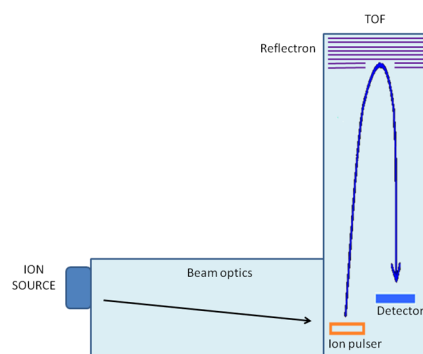


Fig. 3.11. Schematic overview of TOF analyzer.

To ensure the mass accuracy of this equipment both external and internal calibration are important. External calibration is carried out prior to analysis by recording the spectra of known reference compounds and assigning them to their exact masses by instrument data system to produce the correct calibration. On the other hand, internal calibration enables recalibration of the m/z scale using a single ion or a series of ions of known m/z from a reference compound that is introduced into the ion source at the same time as the analytes, while the analysis is being carried out⁴⁷.

✓ Tandem and hybrid mass analyzers

Combining different analyzers in sequence is another trend in mass analyzer development, as versatility is increased and it allows performing multiple experiments. Tandem and hybrid mass analyzers are commonly used for tandem mass spectrometry (MS/MS) applications, where two analyzers are combined in series with a reaction chamber in between and the m/z of defined ions are measured before and after the reaction⁴⁹. Triple-quadrupole (QQQ) or quadrupole-time-of-flight (Q-TOF) are some examples in which the combination of different analyzers make possible collecting mass spectra of the decomposition of an ion selected in the first analyzer⁴⁵.

QQQ is a tandem mass analyzer in which precursor ion selection, collision induced dissociation and mass analysis of the product ions take place in one type of mass analyzer, a quadrupole, in triplicate as showed in Fig. 3.12.

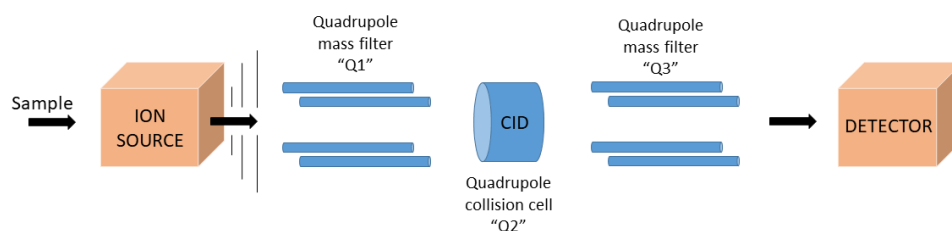


Fig. 3.12. Diagram of a triple quadrupole mass spectrometer.

On the contrary, in hybrid instruments, the stages of mass analysis are carried out in two different types of mass analyzers, like Q-TOF (Fig. 3.13). Q-TOF instrument works as a modified QQQ instrument, where the Q3 quadrupole has been replaced by a TOF mass analyzer. Consequently, when measuring in MS mode, the Q1 quadrupole is operated in radio frequency (RF) only mode and collision cell with low collision energy, whereas in MS/MS mode the Q1 quadrupole selects the precursor ion with unit-mass resolution and is fragmented in the collision cell. Q-TOF analyzers are widely used in structure elucidation, with the advantage over QQQ of performing accurate mass determination (< 5 ppm)⁴⁹.

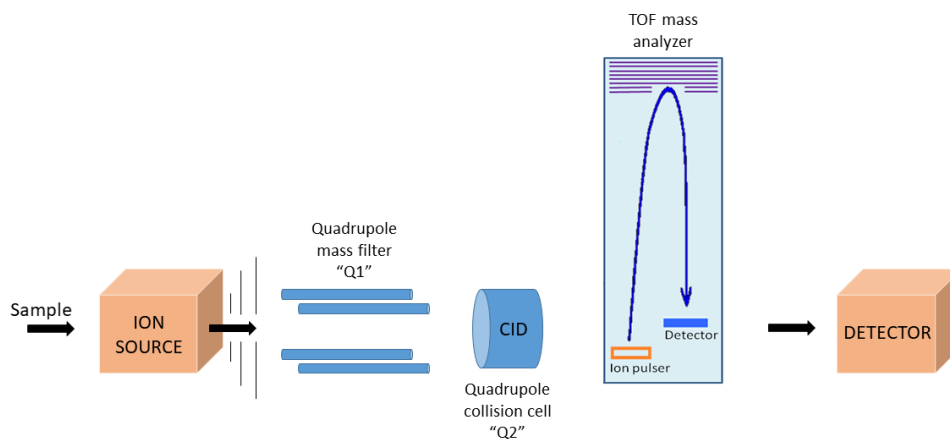


Fig. 3.13. Diagram of a QTOF mass spectrometer.

MS/MS instruments are able to operate in different analysis modes, such as product ion scan, precursor ion scan, neutral-loss scan or multiple reaction monitoring mode (see Fig. 3.14).

In product ion scan Q1 is set to allow the transmission of a unique m/z (parent ion), which collides in Q2 to create fragments or product ions, which are scanned in Q3. Product ion scan is used in screening experiments to obtain the fragmentation patterns of particular compounds.

Precursor ion scan mode consists in letting precursor ions pass through Q1, all the precursor ions collide in Q2 to create product ions, from which only a fragment ion with a defined m/z is allowed to pass through Q3. Precursor ion scans are used in screening experiments where the same fragment ion is expected from a group of compounds.

Regarding multiple reaction monitoring (MRM) mode, a precursor ion is selected in Q1, collides in Q2, and from all the products ion created, only defined m/z passes through Q3. MRM mode is commonly used for analytical methods aimed at quantification in MS/MS.

Finally, in neutral loss scan, precursor ions scanned in Q1 collide in Q2 and Q3 is offset only for those compounds that give a fragment having an specific neutral loss. Neutral loss scans are used for screening experiments with a group of compounds giving the same loss⁴⁹.

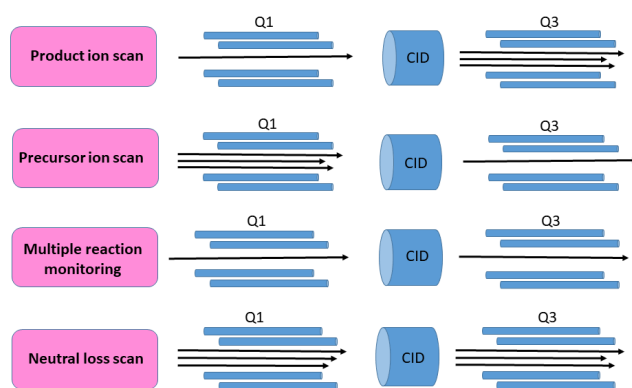


Fig. 3.14. Analysis modes of a MS/MS instrument.

3.3.2.3. Detectors

The ions coming from the mass analyzer are then detected and transformed into a usable signal, that is, an electric current proportional to their abundance⁵⁷. There are different types of ion detection devices, which are backed by fast electronics to enable the high-speed data acquisition required in MS. Electron multipliers are based on the repeated emissions of secondary electrons resulting from the repeated collisions of energetic particles at a convenient surface and is the most widely used detection system. The electron multiplier is used in combination with quadrupole, ion-trap and sector instrument,

whereas TOF instruments usually work with microchannel plate (MCP) detectors instead. MCP detectors are arrays of miniature electron multipliers oriented parallel to one another. In addition, TOF instruments contain an analog-to-digital converters (ADC) or a time-to-digital converter (TDC), in order to generate spectra from ion arrival events⁴⁹.

3.4. STATISTICAL AND CHEMOMETRICAL APPROACH

Metabolomic studies generate high-dimensional and complex datasets that remain challenging to handle and fully exploit⁵⁸. Chemometric tools allow extracting the statistically relevant information, which might remain unseen with the naked eye, to differentiate between control individuals and patients suffering from a disease. Both univariate analysis and multivariate analysis techniques are routinely used to provide biological knowledge on the problem studied⁵⁹.

Univariate methods analyze only one variable at a time and include tests to compare different sets of samples, such as Student's *t*-test or Analysis of Variance (ANOVA)⁵⁹. Targeted and untargeted metabolomics data sets include tenths to hundreds of metabolites, which makes necessary to perform multiple test correcting to protect against the increasing probability of having false positives⁵⁹.

On the contrary, **multivariate methods** handle more than one variable at a time and check for the relations of two to hundreds or thousands of variables simultaneously. The possibility that one metabolite from a biological system contains the maximum information to distinguish between different conditions is limited, and a combination of changes in several metabolites, the so-called metabolite pattern, is more prone to store maximum information. Principal component analysis (PCA), an unsupervised method, and partial least squares discriminant analysis (PLS-DA), a supervised method, are the key methods in metabolomics⁵⁹. In addition, some other more new methods, such as, sparse partial least squares discriminant analysis (SPLS-DA) are being used in metabolomics for variable selection and classification in one step^{60,61}. Besides, hierarchical clustering, an unsupervised method have also been used in metabolomics for unsupervised analysis and outlier detection⁶².

When comparing the results obtained from univariate and multivariate analysis methods, it is often thought that if relevant information is found in a set of metabolites, each of the

individual metabolites also contains information. However, this is not necessarily true, as often information is only present in the relations between metabolites, as a pattern⁵⁹.

Multivariate methods require the use of data pretreatment methods (see Table 3.2), like scaling and/or normalization. For instance, PCA or PLS are maximum variance projection methods and the use of non-normalized variables would provide stronger weights to variables with large variance and would be more likely to be addressed in the constructed model⁶³. There are several scaling and normalization methods which stress different aspects of the experimental data and have both advantages and disadvantages⁶⁴.

Table 3.2. Data pretreatment methods in metabolomics, being \bar{x}_i the mean value for each variable and SD the standard deviation. Extracted from Van den Berg *et al.*¹.

Data pretreatment	Method	Formula	Advantages and disadvantages
Centering	Meancentering	$x'_{ij} = x_{ij} - \bar{x}_i$	<input checked="" type="checkbox"/> Focuses on the differences by removing the offset from the data <input checked="" type="checkbox"/> Not enough if data is heterocedastic
	Autoscaling (=unit variance scaling = Z transform)	$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{SD}$	<input checked="" type="checkbox"/> Compares metabolites based on correlations and all the metabolites become equally important. <input checked="" type="checkbox"/> There is a chance of inflating the measurement errors.
Scaling	Pareto scaling	$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{SD}}$	<input checked="" type="checkbox"/> Reduces the relative importance of large values, while keeping data structure partially intact. It is sensitive to large fold changes. <input checked="" type="checkbox"/> Inflation of the measurement errors.
	Range scaling	$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{(x_{imax} - x_{imin})}$	<input checked="" type="checkbox"/> Compares metabolites taking into account the relative biological response range. All metabolites become equally important. <input checked="" type="checkbox"/> Inflation of the measurement errors and sensitive to outliers.
	Vast scaling	$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{SD} \cdot \frac{\bar{x}_i}{SD}$	<input checked="" type="checkbox"/> Focuses on the metabolites with small fluctuations. Robust. Prior knowledge can be used. <input checked="" type="checkbox"/> Not suited for large variation without group structure.
	Level scaling	$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{x}_i}$	<input checked="" type="checkbox"/> Focuses on relative response. <input checked="" type="checkbox"/> Inflation of the measurement errors.
	Log transformation	$x'_{ij} = \log_{10}(x_{ij})$	<input checked="" type="checkbox"/> Reduces heteroscedasticity and makes multiplicative models additive. <input checked="" type="checkbox"/> Difficulties with values with large SD and zeros.
Transformation	Power transformation	$x'_{ij} = \sqrt{x_{ij}}$	<input checked="" type="checkbox"/> Reduces heteroscedasticity and has no problem with small values. <input checked="" type="checkbox"/> The square root is arbitrary.

3.4.1. HIERARCHICAL CLUSTERING

Natural pattern identification in data is one of the most important goals of chemometrics⁶⁵. Hierarchical clustering is an unsupervised method that makes groups of samples according to the structure of high dimensional data sets, based on object-wise similarities or distances. Indeed, data often contains a hierarchical structure in which groups or clusters contain mutually exclusive sub-groups⁶⁶.

Hierarchical clustering is carried out by means of different algorithms. Single-linkage algorithm is based on taking the shortest distance between clusters. On the contrary, complete linkage clustering is performed using largest distances between clusters, that is, the distances of members of the same cluster must be small. Besides, average linkage clustering is an intermediate approach in which average distance between cluster members is used. Typically, single linkage algorithm leads to large clusters with a few small groups, or even individual samples, whereas complete and average linkage are more prone to provide more compact and rounded groups of samples⁶⁶. Hierarchical clustering is used to find natural groupings of data as well as to identify outliers, data points very different from the rest of the data based on a set of measures⁶⁷.

Dendrograms are graphical representations to visualize the results from hierarchical clustering, which indicate the "distance" between different groups in the y-axis, whereas the connections show the joins found.

3.4.2. PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA searches for common patterns in a dataset by establishing new directions and compressing the dataset into a lower number of variables, also known as principal components (PC) that maximize the variance. Using PCs a new coordinate system can be constructed where the data space is projected so that a new coordinate axis is built in the direction of the largest variance. Scores are the projection of the data onto this new coordinate system and loadings define the contribution of each original variable or metabolite to the components. Even if PCA is an unsupervised analysis method, the *a priori* knowledge can be used to colour objects in the scoreplot emphasizing potential patterns and/or quantitative gradients found in data⁶⁴.

Singular Value Decomposition (SVD) is a fast and numerically stable algorithm used to calculate principal components (see Equation 3.6). It works decomposing original data matrix (X) into three parts: an orthonormal matrix containing left singular vectors (U), a diagonal matrix containing singular values (D) and an orthonormal matrix containing right singular vectors (P). Superscript T means the transposition of a matrix⁶⁸.

$$X = (UD)P^T = TP^T$$

Eq. 3.6.

Score matrix (T) is obtained from the multiplication of U and D matrices and constitutes the coordinates in the space of the PC, whereas columns in P matrix represents loadings. The variance in X is the sum of the squared singular values, which represent the variance explained by each PC. PCs are ordered by explained variation, i.e. the first PC explains highest variation, followed by the second, the third, etc. No relevant information is found from a certain PC upwards and once selected the appropriate number of PCs, it is possible to build a reconstruction (\tilde{X}) of the original data matrix (X) up to the selected number of PCs, being T and P score and loading matrices and matrix E the residuals, as showed in Equation 3.7.

$$X = TP^T = \tilde{X} + E$$

Eq. 3.7.

Different strategies can be followed in order to assess the appropriateness of the chosen number of PCs. On the one hand, it is possible to plot the cumulative fraction of variance explained and set an arbitrary cut-off value, such as 80 or 90 % of variance. Variance calculation is showed in Equation 3.8. On the other hand, a Scree plot might be represented, which depicts the logarithm of variance explained versus PC number. The number of PCs for matrix reconstruction in Scree plots is selected according to a characteristic elbow or slope change showed graphically.

$$\text{Var}(\text{PC}_i) = \frac{d_i^2}{\sum_{j=1} d_j^2}$$

Eq. 3.8.

3.4.3. CLASSIFICATION METHODS

Despite the fact that PCA is a very powerful explorative method, supervised classification methods, such as PLS-DA, are preferred. It is necessary to split the data randomly into training and test sets for classification error estimation. Indeed, if any information from test set is included into the model, error estimation would be biased and the results would be too optimistic. In addition, any scaling method must be performed after splitting data into training and test sets, and the test set needs to be scaled based on means and standard deviations from the training set.

Expected classification performance or error of prediction should be obtained using cross validation or similar techniques. Cross validation deals with the idea of dividing data into different training and test sets, to avoid a performance dependent only in the random division of data. Cross validation works leaving out a certain part of data to be used as test set and the rest of data is used to build the model as a training set. After estimating the performance of the model, data set is again divided into new training and test set, leaving out a different part of data. This process is repeated once and again until all samples have been left out once and average performances are obtained (see Fig. 3.15). Different cross validation strategies can be followed. When size of the test set is one, this procedure is known as Leave-one-out (LOO) cross-validation. It is also possible to leave a larger fraction of data, by using a percentage of total data.

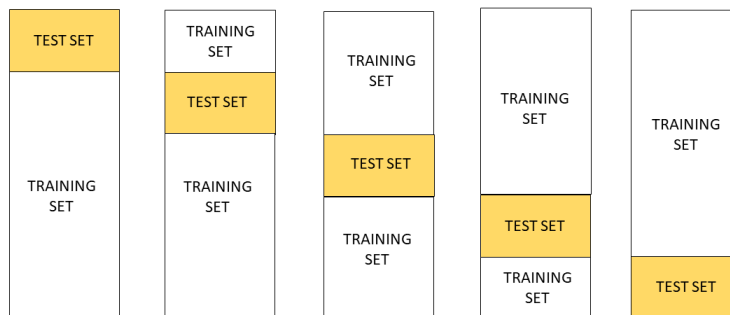
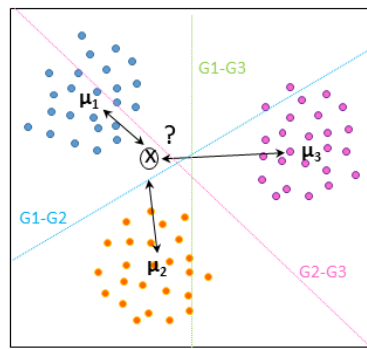


Fig. 3.15. Division of data set into training sets and test sets for cross validation.

3.4.3.1. Traditional discrimination analysis methods

A set of techniques are recognized under the header traditional discrimination analysis methods: maximum likelihood linear discriminant analysis (ML-LDA), Fisher linear discriminant analysis (Fisher-LDA), quadratic discriminant analysis (QDA) and *k*-nearest neighbours (*k*NN)⁶⁹.

ML-LDA consists on finding the cluster centres of each class and according to these midpoints, drawing separating hyperplanes between these centres. Samples are assigned to the class which contains the closest midpoint. The distance between samples and midpoints is measured according to the squared Mahalanobis distance. In this case, covariances (Σ) needed for Mahalanobis are calculated using a pooled covariance matrix estimated from data by pooling the individual sample covariance matrices (S_i).



Pooled covariance matrix calculation:

$$S = \frac{1}{n - G} \sum_{i=1}^G n_i S_i$$

- G: groups
- n_i : number of samples in group *i*
- N: total number of samples
- S_i : individual sample covariance matrices

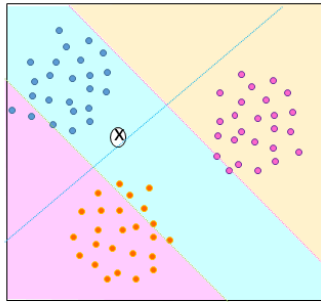
Mahalanobis distance calculation:

$$\text{Mahalanobis distance} = (x - \hat{\mu}_i) \hat{\Sigma}^{-1} (x - \hat{\mu}_i)^T$$

- $\hat{\mu}_i$: centre of a group
- x*: test sample
- $\hat{\Sigma}$: pooled covariance matrix (*S* in ML-LDA)

Fig. 3.16. An example of the calculation of a new sample according to its Mahalanobis distance to the centres of three different classes using ML-LDA.

Regarding Fisher-LDA, it is similar to ML-LDA with the difference of finding linear combination of variables that maximizes the ratio of within-group sums of squares and between-group sum of squares. That is, this classification method finds a linear combination that leads to small distances within a class and large distances between classes. It should be noted that for two-group cases the same solution as for ML-LDA is achieved, whereas for more than two groups, the results will differ unless sample means are collinear.



Fisher criterion:

$$J(a) = \frac{a^T B a}{a^T W a}$$

B: between-group covariance matrices

W: within-group covariance matrices

$$B = \sum_{i=1}^G n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

$$W = \sum_{i=1}^G \tilde{X}_i^T \tilde{X}_i$$

\tilde{X}_i : centred data matrix containing samples of class i

\bar{x}_i : mean vector for class i

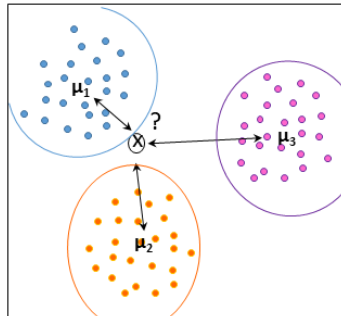
\bar{x} : mean vector for the whole data matrix

Solving the equation:

$$J(a) = W^{-1}B$$

Fig. 3.17. Summary of the algorithm used for Fisher-LDA classification.

On the other hand, QDA works similar to ML-LDA with the difference that instead of a pooled covariance matrix, each class is described by its own covariance matrix.



Covariance matrix calculation per group:

$$S = \frac{1}{n-1} (X)^T (X)$$

(X): meancentered matrix

n: number of samples

Mahalanobis distance calculation:

$$\text{Mahalanobis distance} = (x - \hat{\mu}_i) \hat{\Sigma}_i^{-1} (x - \hat{\mu}_i)^T$$

$\hat{\mu}$: centre of a group

x: test sample

$\hat{\Sigma}$: covariance matrix of each group (S in QDA)

Fig. 3.18. Calculation of the classification of a new sample according to its Mahalanobis distance to the centres of three different classes using QDA.

Concerning k NN, this technique is based on a completely different approach; samples are classified according to the class of the closest k number of samples following the majority vote. However, its main disadvantage is that in principle the whole training set comprises the model and should be saved for the predictions for new objects. This can be slow and may present memory problems, unless objects are removed to simplify the model.

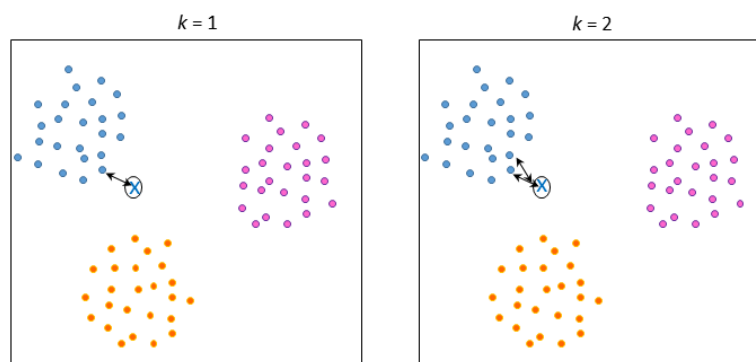


Fig. 3.19. An example of the classification of a new sample according to $k=1$ and $k=2$ nearest neighbours.

3.4.3.2. Partial least squares discriminant analysis (PLS-DA)

PLS-DA method uses a classifier variable (typically 0 and 1, or -1 and 1 values) to describe the condition of subjects and predict it using the information found in the metabolite data set. PLS-DA classification method is based on partial least-squares regression (PLS). PLS does regression using latent variables (LV), which means that the number of LV to be used in a PLS-DA model should be optimized on the basis of classification performance, that is, the rate of correctly classified samples⁶⁴.

PLS-DA classification is based on PLS regression algorithm (see Eq. 3.9) and needs for at least meancentering data. X is used to define scaled data matrix, whereas Y is the matrix with the values to be predicted. Both, X and Y matrices contain score (T for X -scores and U for Y -scores) and loading matrices (P for X -loadings and Q for Y -loadings). The regression model is built using the reconstructed matrix (see Eq. 3.7) and during each iteration the variation associated with the estimated component is removed from data and the remainder (E for the deflated X matrix and F for the deflated Y matrix) is used to estimate the next latent variable⁶⁸.

PLS ALGORITHM

1. Cross product matrix is first obtained: $S = E' \cdot F$
2. Singular value decomposition is performed from S and the so-called weights (w) are obtained and stored in different columns of W matrix.
3. X-scores are obtained: $t = E \cdot w$
4. X-loadings are calculated: $P = \frac{E' \cdot t}{(t' \cdot t)}$
5. Y-loadings are obtained: $Q = \frac{F' \cdot t}{(t' \cdot t)}$
6. Deflated matrices are obtained and used during next iteration (until a defined number of LV or until all LV have been calculated):

$$E = E - t \cdot p'$$

$$F = F - t \cdot q'$$
7. Y-scores are calculated: $U = Y \cdot Q$
8. A different representation of weights is used for the estimation of regression coefficients (B):

$$R = W \cdot \text{inv}(P' \cdot W)$$

$$A = \text{inv}(T' \cdot T) * T' \cdot Y$$

$$\text{Bmat} = R \cdot A$$

$$b_0 = \text{mean}(Y) - \text{mean}(X \cdot \text{Bmat})$$

$$B = [b_0; \text{Bmat}]$$
9. Prediction is carried out: $YP = X \cdot \text{Bmat} + b_0$

** Note that capital letters refer to the whole matrix and lower-case letters to the part of the matrix corresponding to a defined latent variable.*

Eq. 3.9.

Once predicted Y matrix, discriminant analysis is performed and samples are classified to a defined group according to the established threshold. For instance, if -1 value has been used for group A and 1 value for group B, the threshold to classify the samples would be 0, and all the samples below 0 would be classified in group A, whereas samples above 0 would correspond to group B.

3.4.4. SPARSE PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS (SPLS-DA)

SPLS-DA performs variable selection in the X data set with Lasso penalization and classification based on supervised framework in one step, i.e. the selection of a defined number of variables is performed according to the different classes of the samples⁷⁰.

SPLS-DA is used after optimizing two parameters: the number of latent variables (H) and the number of selected variables for each latent variables⁶¹. The number of latent variables $H=K-1$ is usually advised in multiclass problems with K classes to obtain the most stable models⁷⁰. Regarding the number of selected variables, it depends on the

complexity of data and it should be optimized taking into account model performance and stability⁶¹.

To sum up, the improvement of technology has enabled the development of metabolomics as an analytical tool, which is an important source of new biomarkers for a variety of diseases. However, to be successful at biomarker discovery, careful planning of experiments, use of analytical instruments and statistical and chemometric analyses are essential⁷¹.

3.5. BIBLIOGRAPHY

1. Van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*. 2006; 7.
2. Hasin Y, Seldin M, Lusi A, Hasin Y, Lusi A, Lusi A. Multi-omics approaches to disease. *Genome Biol*. 2017; 18: 83.
3. Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynecol*. 2011; 13: 189-195.
4. Fiehn O. Metabolomics - the link between genotypes and phenotypes. *Plant Mol Biol*. 2002; 48: 155-171.
5. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2012; 883-884: 68-75.
6. Monteiro MS, Carvalho M, Bastos ML, Guedes de Pinho P. Metabolomics analysis for biomarker discovery: advances and challenges. *Curr Med Chem*. 2013; 20: 257-271.
7. Patti GJ, Yanes O, Siuzdak G. Innovation Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*. 2012; 13: 263-269.
8. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. *TrAC-Trend Anal Chem*. 2005; 24: 285-294.
9. German JB, Gillies LA, Smilowitz JT, Zivkovic AM, Watkins SM. Lipidomics and lipid profiling in metabolomics. *Curr Opin Lipidol*. 2007; 18: 66-71.
10. Nicholson JK, Lindon JC, Holmes E. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. 1999; 29: 1181-1189.
11. Barderas MG, Laborde CM, Posada M, de ICF, Zubiri I, Vivanco F, et al. Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases. *J Biomed Biotechnol*. 2011; 2011: 790132.

12. Deo RC, Hunter L, Lewis GD, Pare G, Vasan RS, Chasman D, et al. Interpreting metabolomic profiles using unbiased pathway models. *PLoS Comput Biol.* 2010; 6: No pp. given.
13. Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG. Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol.* 2005; 3: 557-565.
14. Pope GA, MacKenzie DA, Defemez M, Aroso MAMM, Fuller LJ, Mellon FA, et al. Metabolic footprinting as a tool for discriminating between brewing yeasts. *Yeast.* 2007; 24: 667-679.
15. Llorach R, Garcia-Aloy M, Tulipani S, Vazquez-Fresno R, Andres-Lacueva C. Nutrimetabolomic Strategies To Develop New Biomarkers of Intake and Health Effects. *J Agric Food Chem.* 2012; 60: 8797-8808.
16. Dudley E, Yousef M, Wang Y, Griffiths WJ. Targeted metabolomics and mass spectrometry. *Adv Protein Chem Struct Biol.* 2010; 80: 45-83.
17. Theodoridis G, Gika HG, Wilson ID. Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. *Mass Spectrom Rev.* 2011; 30: 884-906.
18. Wishart DS. Metabolomics: the principles and potential applications to transplantation. *Am J Transplant.* 2005; 5: 2814-2820.
19. Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Anal Chim Acta.* 2012; 750: 82-97.
20. Bothwell JHF, Griffin JL. An introduction to biological nuclear magnetic resonance spectroscopy. *Biol Rev Camb Philos Soc.* 2011; 86: 493-510.
21. Markley JL, Bruschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, et al. The future of NMR-based metabolomics. *Curr Opin Biotechnol.* 2017; 43: 34-40.
22. Alonso A, Marsal S, Julia A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol.* 2015; 3: 23.
23. Larive CK, Barding GA, Dinges MM. NMR Spectroscopy for Metabolomics and Metabolic Profiling. *Anal Chem.* 2015; 87: 133-146.
24. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. *Analyst.* 2012; 137: 293-300.
25. Boccard J, Veuthey J-L, Rudaz S. Knowledge discovery in metabolomics: an overview of MS data handling. *J Sep Sci.* 2010; 33: 290-304.
26. De Hoffmann E, Stroobant V. Introduction. In: de Hoffmann E, Stroobant V, editors. *Mass spectrometry: Principles and applications.* 3rd ed, 2007; p. 1-13.
27. Ho CS, Lam CWK, Chan MHM, Cheung RCK, Law LK, Lit LCW, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev.* 2003; 24: 3-12.

28. Skoog DA, Holler FJ, Crouch SR. Cromatografía de líquidos. In: learning C, editor. Principios de análisis instrumental. 6th ed, 2008; p. 816-817.
29. Zhou B, Xiao JF, Tuli L, Ransom HW. LC-MS-based metabolomics. *Mol BioSyst.* 2012; 8: 470-481.
30. Ranjbar MRN, Luo Y, Di Poto C, Varghese RS, Ferrarini A, Zhang C, et al. GC-MS based plasma metabolomics for identification of candidate biomarkers for hepatocellular carcinoma in Egyptian cohort. *PLoS One.* 2015; 10.
31. Ibañez C, Simo C, Garcia-Canas V, Gomez-Martinez A, Ferragut JA, Cifuentes A. CE/LC-MS multiplatform for broad metabolomic analysis of dietary polyphenols effect on colon cancer cells proliferation. *Electrophoresis.* 2012; 33: 2328-2336.
32. Lemièrre F. Interfaces for LC-MS. *LC GC Eur.* 2001; *Guide to LC-MS:* 2-8.
33. Wilm M. Principles of electrospray ionization. *Mol Cell Proteomics.* 2011; 10.
34. De Hoffmann E, Stroobant V. Ion sources. In: de Hoffmann E, Stroobant V, editors. *Mass Spectrometry: Principles and Applications* 3rd ed, 2007; p. 15-83.
35. Cech NB, Enke CG. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev.* 2002; 20: 362-387.
36. Banerjee S, Mazumdar S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int J Anal Chem.* 2012; 2012: 40.
37. Lavagnini IM, Franco; Seraglia, Roberta; Traldi, Pietro. *Quantitative Applications of Mass Spectrometry* John Wiley & sons, LTD; 2006.
38. Górecki T. Chromatography. In: Nolle LML, editor. *Chromatographic Analysis of the Environment.* 93. 3rd ed: CRC Press; 2005; p. 133-176.
39. Balogh MP. Source design and the utility of multimode ionization. *Spectroscopy.* 2004; 19: 52-56.
40. Cai S-S, Syage JA. Comparison of Atmospheric Pressure Photoionization, Atmospheric Pressure Chemical Ionization, and Electrospray Ionization Mass Spectrometry for Analysis of Lipids. *Anal Chem.* 2006; 78: 1191-1199.
41. Greig MJ, Bolanos B, Quenzer T, Bylund JMR. Fourier transform ion cyclotron resonance mass spectrometry using atmospheric pressure photoionization for high-resolution analyses of corticosteroids. *Rapid Commun Mass Spectrom.* 2003; 17: 2763-2768.
42. Stump MJ, Fleming RC, Gong W-H, Jaber AJ, Jones JJ, Surber CW, et al. Matrix-assisted laser desorption mass spectrometry. *Appl Spectrosc Rev.* 2002; 37: 275-303.
43. Dreisewerd K. The desorption process in MALDI. *Chem Rev.* 2003; 103: 395-425.
44. Fischer SM, Perkins PD. Agilent Technologies. Technical Overview: Simultaneous Electrospray and Atmospheric Pressure Chemical Ionization: The Science Behind the Agilent Multimode Ion Source. 2005; 5989-2935EN: 1-20.

45. De Hoffmann E, Stroobant V. Mass Analysers. In: de Hoffmann E, Stroobant V, editors. *Mass Spectrometry: Principles and Applications* 3rd ed, 2007; p. 85-173.
46. Holcapek M, Jirasko R, Lisa M. Recent developments in liquid chromatography-mass spectrometry and related techniques. *J Chromatogr A*. 2012; 1259: 3-15.
47. Webb K, Bristow T, Sargent M, Stein B. *Methodology for Accurate Mass Measurement of Small Molecules*. Best Practice Guide. LGC Limited. 2004.
48. Brenton AG, Godfrey AR. Accurate Mass Measurement: Terminology and Treatment of Data. *J Am Soc Mass Spectrom*. 2010; 21: 1821-1835.
49. Niessen WMA, Falck D. Introduction to Mass Spectrometry, a Tutorial. In: Kool J, Niessen WMA, editors. *Analyzing Biomolecular Interactions by Mass Spectrometry*. 2015; p. 1-54.
50. Bakalarski CE, Haas W, Dephoure NE, Gygi SP. The effects of mass accuracy, data acquisition speed, and search algorithm choice on peptide identification rates in phosphoproteomics. *Anal Bioanal Chem*. 2007; 389: 1409-1419.
51. Smedsgaard J. Analytical Tools. In: Villas-Bôas SG, Roessner U, Hansen MAE, Smedsgaard J, Nielsen J, editors. *Metabolome Analysis: An introduction*. Wiley; 2007; p. 83-145.
52. Stephens WE. A pulsed mass spectrometer with time dispersion. *Phys Rev*. 1946; 69: 691.
53. Wiley WC, McLaren IH. Time-of-flight mass spectrometer with improved resolution. *Rev Sci Instrum*. 1955; 26: 1150-1157.
54. Guilhaus M, Selby D, Mlynski V. Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom Rev*. 2000; 19: 65-107.
55. Sriveena T, Srividya A, Ajitha A, Rao VUM. Time of flight mass spectrometry: review. *World J Pharm Pharm Sci*. 2015; 4: 614-625.
56. Lacorte S, Fernandez-Alba AR. Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food. *Mass Spectrom Rev*. 2006; 25: 866-880.
57. De Hoffmann E, Stroobant V. Detectors and computers. In: de Hoffmann E, Stroobant V, editors. *Mass Spectrometry: Principles and Applications* 2007; p. 175-188.
58. Tugizimana F, Steenkamp PA, Piater LA, Dubery IA. A Conversation on Data Mining Strategies in LC-MS Untargeted Metabolomics: Pre-Processing and Pre-Treatment Steps. *Metabolites*. 2016; 6.
59. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*. 2014; 10: 361-374.
60. Jiang M, Wang C, Zhang Y, Feng Y, Wang Y, Zhu Y. Sparse Partial-least-squares Discriminant Analysis for Different Geographical Origins of *Salvia miltiorrhiza* by 1H-NMR-based Metabolomics. *Phytochem Anal*. 2014; 25: 50-58.

61. Szymanska E, Brodrick E, Williams M, Davies AN, van Manen H-J, Buydens LMC. Data Size Reduction Strategy for the Classification of Breath and Air Samples Using Multicapillary Column-Ion Mobility Spectrometry. *Anal Chem.* 2015; 87: 869-875.
62. Madala NE, Piater LA, Steenkamp PA, Dubery IA. Multivariate statistical models of metabolomic data reveals different metabolite distribution patterns in isonitrosoacetophenone-elicited *Nicotiana tabacum* and *Sorghum bicolor* cells. *SpringerPlus.* 2014; 3: 254.
63. Bouhifd M, Hartung T, Hogberg HT, Kleensang A, Zhao L. Review: Toxicometabolomics. *J Appl Toxicol.* 2013; 33: 1365-1383.
64. Savorani F, Rasmussen MA, Mikkelsen MS, Engelsen SB. A primer to nutritional metabolomics by NMR spectroscopy and chemometrics. *Food Res Int.* 2013; 54: 1131-1145.
65. Almeida JAS, Barbosa LMS, Pais AACC, Formosinho SJ. Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering. *Chemom Intell Lab Syst.* 2007; 87: 208-217.
66. Cluster analysis. In: Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J, editors. *Handbook of Chemometrics and Qualimetrics: Part B.* 1998; p. 57-86.
67. Haddad I, Hiller K, Frimmersdorf E, Benkert B, Schomburg D, Jahn D. An emergent self-organizing map based analysis pipeline for comparative metabolome studies. *In Silico Biol.* 2009; 9: 163-178.
68. Analysis of Measurement Tables. In: Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J, editors. *Handbook of Chemometrics and Qualimetrics: Part B.* 1998; p. 87-160.
69. Supervised pattern recognition. In: Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J, editors. *Handbook of Chemometrics and Qualimetrics: Part B.* 1998; p. 207-241.
70. Le Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics.* 2011; 12: 253.
71. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics. A review in human disease diagnosis. *Anal Chim Acta.* 2010; 659: 23-33.



CHAPTER IV.

Targeted metabolomics analysis of amino acids and related compounds using LC-QTOF-MS

- 4.1. Introduction**
- 4.2. Metabolic pathways suspicious of being affected in paediatrics with CKD: literature review**
- 4.3. Targeted metabolomics workflow**
- 4.4. Study's goal**
- 4.5. Description of targeted metabolites**
- 4.6. Materials and equipment**
 - 4.6.1. Standards and reagents**
 - 4.6.2. Instrumental analysis**
 - 4.6.3. Preparation of standard solutions**
 - 4.6.4. Subjects and sampling**
- 4.7. Optimization of the analytical method**
 - 4.7.1. Selection of the chromatographic variables**
 - 4.7.2. Selection of mass spectrometry variables**
 - 4.7.3. Optimization of sample treatment conditions**
- 4.8. Analytical method evaluation**
 - 4.8.1. Calibration of the method and matrix effect**
 - 4.8.2. Limits of quantification, accuracy and precision**
 - 4.8.3. Recovery from the extraction process**
 - 4.8.4. Stability**
- 4.9. Sample analysis**
- 4.10. Data analysis**
 - 4.10.1. Descriptive statistics**
 - 4.10.2. Univariate analysis**
 - 4.10.3. Multivariate analysis**
- 4.11. Biological interpretation of the results**
- 4.12. Conclusions**
- 4.13. Bibliography**

4.1. INTRODUCTION

CKD is characterized by a progressive loss of renal function over a period of months or years that causes the accumulation of potentially toxic compounds. These compounds, also called uremic toxins, adversely affect biological functions and were collected by the European Uremic Toxin Work Group (*EUTox*) of the European Society for Artificial Organs in 2003 for the first time^{1,2} and updated last in 2012³. The accumulation of uremic toxins is in relation with the alteration of different metabolic pathways, which may lead to an increase or decrease in concentration of some compounds⁴⁻⁶.

Amino acids are known to participate in several metabolic pathways (e.g.: urea cycle⁷, arginine methylation⁸, citric acid cycle⁹, arginine-creatine¹⁰, amino acid biosynthesis¹¹ or carbon metabolism¹²). Prior knowledge from literature shows that some amino acids, amino acid derivatives and related compounds from different metabolic pathways could be potential targets for CKD diagnosis in paediatrics⁸.

There are 21 proteinogenic amino acids in eukaryotes, also called standard or classical amino acids, which are genetically encoded and are incorporated into proteins. Arginine, cysteine, glycine and methionine are some of these amino acids. Nonstandard amino acids are produced by metabolic conversions of free amino acids (e.g., citrulline), or by post-translational modification (hydroxylation, methylation or carboxylation) of amino acids present in proteins¹³. Amino acids might also include new functional groups to become amino acid derivatives, such as, dimethylglycine, betaine, homocysteine, homoarginine, symmetric dimethylarginine, asymmetric dimethylarginine, S-adenosylmethionine and S-adenosylhomocysteine.

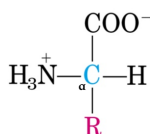


Fig. 4.1. Typical structure of α -amino acids.

Amino acids present in proteins contain one carboxyl group, one amino group, one hydrogen and a variable side chain (R) at a single α -carbon atom. Therefore, C_α -atom is asymmetric in all these amino acids except for glycine, where the side chain is an hydrogen. Even if both L- and D-configurations are possible, almost all of the proteinogenic amino acids in nature have the L-configuration, with the exception of D-amino acids found in bacterial envelopes and in some antibiotics^{13,14}.

4.2. METABOLIC PATHWAYS SUSPICIOUS OF BEING AFFECTED IN PAEDIATRICS WITH CKD: LITERATURE REVIEW

Urea cycle, arginine methylation and arginine-creatine metabolic pathways are interconnected by arginine amino acid. A vast amount of amino acids and amino acid derivatives take part in these metabolic routes (see Fig. 4.2).

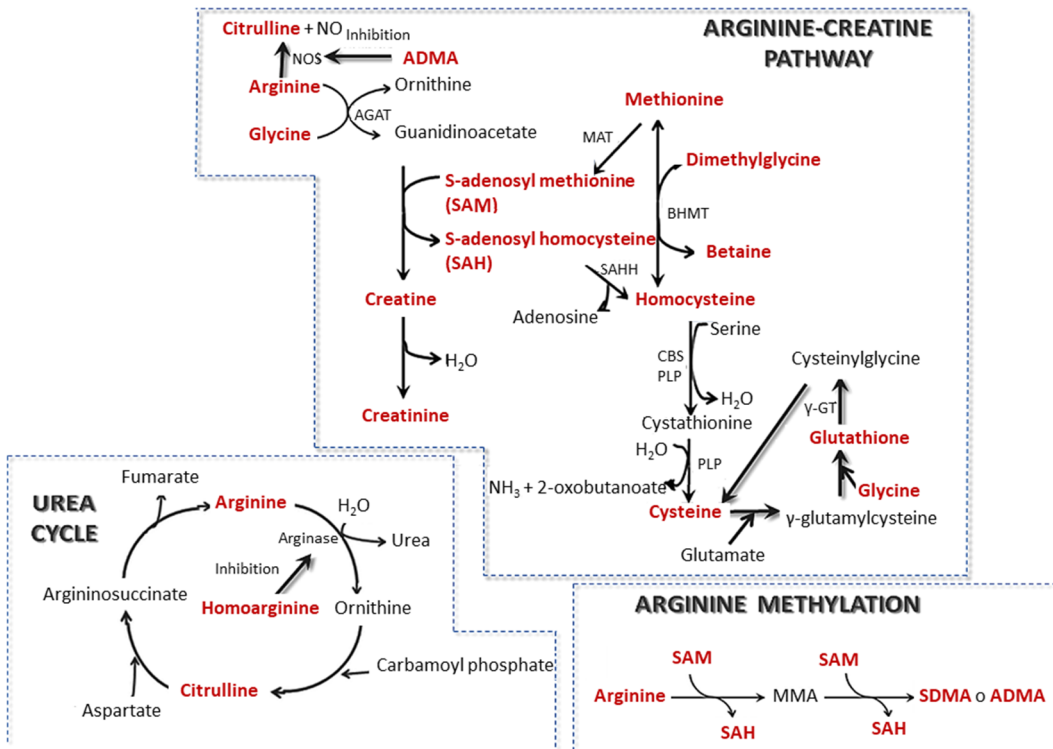


Fig. 4.2. Arginine-creatine metabolic pathway, urea cycle and arginine methylation. AGAT: arginine-glycine amidinotransferase, NOS: nitric oxide sintase, MAT: methionine adenosyltransferase, SAHH: SAH hydrolase, BHMT: betaine-homocysteine S-methyltransferase, PLP: pyridoxal phosphate, CBS: cystathionine beta synthase, γ-GT: gamma glutamil transpeptidase.

When amino groups are not reused for the synthesis of new amino acids or other nitrogenous products, they are channeled into an excretory end product by means of urea cycle. Humans are ureotelic organisms, i.e. ammonia deposited in the mitochondria of hepatocytes is converted into urea, a less toxic substance, within the urea cycle. Afterwards, urea is conducted to the bloodstream and is finally excreted into urine by the kidneys¹⁵.

On the other hand, even though the primary sequence of proteins is given by the genetic code, functional diversity of proteins is achieved by means of different post-translational

modifications, such as phosphorylation or methylation. The nitrogen of arginine within polypeptides can receive methyl groups in a process termed arginine methylation, acting as protein regulators¹⁶.

Regarding arginine-creatine metabolic pathway, it aims at the synthesis of creatine, having the creatine/creatine-phosphate system an important role in the transmission and storage of phosphate-bound energy¹⁷.

An in-depth study on the literature of the most meaningful metabolites of these metabolic routes has been carried out to evaluate the hypothesis that these interconnected metabolic pathways might be affected in paediatrics with CKD. Relevant research shows that several compounds from these interconnected pathways, including creatinine, may be up- or down-regulated. Evidences for increased or decreased levels of the following compounds in addition to creatinine (CNN) are listed in Table 4.2: Asymmetric dimethylarginine (ADMA), arginine (ARG), betaine (BET), citrulline (CIT), creatine (CTN), cysteine (CYS), dimethylglycine (DMG), glutathione (GSH), glycine (GLY), homoarginine (HARG), homocysteine (HCYS), methionine (MET), S-adenosyl homocysteine (SAH), S-adenosyl methionine (SAM) and symmetric dimethylarginine (SDMA).

Table 4.1. Evidences of increased or decreased plasmatic levels of the compounds taking part in the arginine-creatine pathway, arginine methylation and urea cycle.

	Population	Blood concentrations in CKD	Author	Year
ADMA	Paediatrics	Increased in transplanted patients	Andrade ⁸	2011
	Adults	Increased, being correlated with serum creatinine and GFR	Fliser ¹⁸	2005
	Adults	Significantly elevated in all patients comparing to healthy controls	Fleck ¹⁹	2003
	Adults	Increased (doubled levels in ESRD)	Marescau ²⁰	1997
ARG	Paediatrics	Low ARG/ADMA coefficients	Andrade ⁸	2011
	Sprague-Dawley	Synthesis decreased in kidney (could be compensated in other organs)	Chen ²¹	2010
	Adults	Decreased in dependence on the degree of kidney disease	Fleck ¹⁹	2003
BET	Adults	Increased in uremic patients	Mutsaers ²²	2013
	Adults	Decreased in renal disease	Lever ²³	1994
CIT	Sprague-Dawley rats	Moderately increased	Chen ²¹	2010
CTN	Adults	Increased in uremic patients in treatment with dialysis	De Deyn ²⁴	1986
CYS	Adults	Both total and free cysteine increased in both CKD patients receiving and not receiving RRT	Suliman ²⁵	1997
DMG	Adults	Increased in 3-4 CKD stage in comparison with controls	Mutsaers ²²	2013
	Adults	Increased as renal function declined. 2-fold to 3-fold elevated in dialysis patients	McGregor ²⁶	2001
GSH	Paediatrics	Decreased in children on peritoneal dialysis in comparison with a control children group	Zwolinska ²⁷	2009
GLY	Paediatrics	Increased in transplanted children	Andrade ⁸	2011
	Adults	Increased in pre-hemodialysis and in those patients treated with CAPD	Chuang ²⁸	2006
	Male Wistar rats	Slightly increased in CKD with and without acidosis	Holecek ²⁹	2001
HARG	Adults	Low HARG concentration might be an early indicator of CKD	Drechsler ³⁰	2013
	Adults	HARG declines with advancing renal disease	Ravani ³¹	2013
	Adults	Decreased	Marescau ²⁰	1997
HCYS	Paediatrics	HCYS was increased in transplanted children and in children underlying CKD	Andrade ⁸ Andrade ²	2011 2008
	Paediatrics	Total plasma HCYS was increased	Merouani ³²	2001
	Paediatrics	Both total and free HCYS increased in both CKD receiving and not receiving RRT	Suliman ²⁵	1997
MET	Paediatrics	Decreased in haemodialyzed patients in comparison with a control group	El Sawy ³³	2012
	Male Wistar rats	Decreased in those suffering from acidosis and not suffering from acidosis in comparison with controls	Holecek ²⁹	2001
	Adults	Decreased in the CAPD and haemodialyzed patients and tended to be low in nondialyzed uremic patients	Suliman ²⁵	1997
SAH	Paediatrics	Increased in transplanted children	Andrade ⁸	2011
	Adults	Increased	Loehrer ³⁴	1998
	Adults	Increased	Perna ³⁵	1996
SAM	Adults	SAM increased and SAM/SAH methylation index decreased in ESRD patients	Loehrer ³⁴	1998
SDMA	Adults	Increased, correlates well with serum CNN and GFR	Fliser ¹⁸	2005
	Adults	Significantly elevated in all patients (receiving and not receiving RRT) in comparison with healthy controls	Fleck ¹⁹	2003

4.3. TARGETED METABOLOMICS WORKFLOW

Targeted metabolomics refers to a strategy in which a previously defined list of metabolites is measured motivated by specific biochemical questions or hypotheses^{36,37}. These pre-defined metabolites can be associated with specific pathways, enzyme systems or a complete metabolite class³⁸. As the chemical properties of the selected targeted compounds are previously known, specific analytical methods and sample clean-up processes are focused on these properties. This often allows a significant reduction of matrix effects and interferences from accompanying compounds, thus calling for extreme sensitivity^{39,40}.

Analytical strategies for targeted metabolomics include the quantitative analysis of defined metabolites both in controls and in patients, being triple quadrupole mass spectrometry (QQQ) mainly applied for this purpose. Hybrid instruments, in which the third quadrupole is replaced by a time-of-flight (QTOF) or linear ion trap (QTRAP) are also used to perform structural elucidation as well as for quantification purposes^{41,42}. In addition, a liquid chromatography (LC) or gas chromatography (GC) separation step prior to the mass spectrometry detection is highly recommended in order to reduce matrix related ion suppression and to be able to quantify separately isobaric and isomeric compounds³⁸.

Targeted metabolomics workflow usually requires the following steps: analytical method development using standards, sample preparation, validation of the analytical method, real sample treatment and analysis, quantification of the analytes, data analysis and interpretation of the results. Instrumental analysis, sample preparation and data analysis should be optimized calling for extreme sensitivity in the quantification of the targeted metabolites and for a proper biological interpretation of the results.

An example of a targeted metabolomics workflow is showed in Figure 4.3.

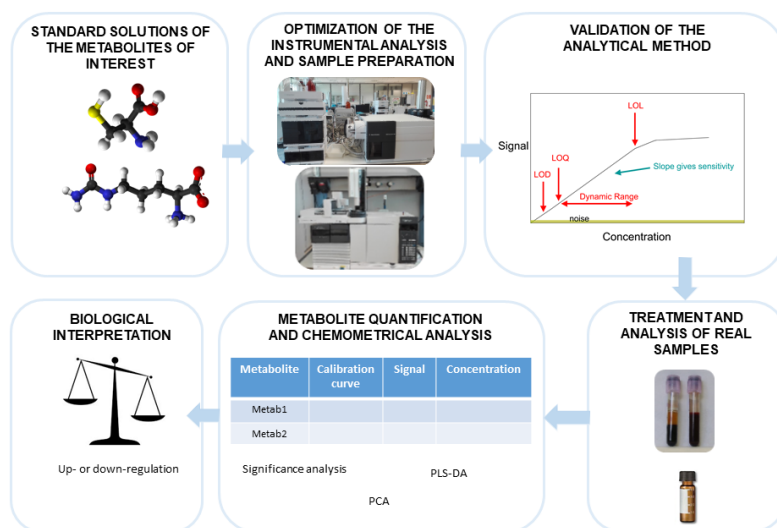


Fig. 4.3. Targeted metabolomics workflow.

4.4. STUDY'S GOAL

The main objective of this study deals with finding amino acids, amino acid derivatives and related compounds suspicious of being altered in paediatrics with CKD and that might be used as effective biomarkers for the early diagnosis. To accomplish this goal, a targeted metabolomics analytical method needs to be developed using LC-MS equipment. Afterwards, this method will be applied to plasma samples from paediatrics suffering from CKD and controls, and data analysis will be necessary to evaluate the results and identify potential biomarkers.

This targeted metabolomics method will focus on the determination and quantification of 16 metabolites: ADMA, ARG, BET, CIT, CTN, CNN, CYS, DMG, GSH, GLY, HARG, HCYS, MET, SAH, SAM and SDMA. All of these compounds are involved in the arginine-creatine metabolic pathway or to related metabolic pathways, such as the urea cycle or methylation of arginine (see Fig. 4.2). L-configuration amino acids will be selected for method development, so unless explicitly stating the opposite, when omitting the information about configuration, L-configuration is implied in this manuscript.

For the development of an analytical method for quantification of these metabolites in plasma, different analytical techniques have been considered: GC-MS, HILIC-MS and ion-pairing-reverse-phase LC-MS. GC-MS was discarded as there is a need for derivatization, reagent elimination and solvent interchange being plasma an aqueous matrix⁴³. In

addition, using HILIC-MS the simultaneous analysis of polar and non-polar compounds is sometimes difficult, and limited stability and column bleeding have been reported⁴⁴⁻⁴⁶. On the contrary, the use of an ion pairing reagent is advantageous as it allows simultaneous determination of polar and non-polar compounds. Besides, using ion pairing reagent instead of derivatization, errors introduced by derivative instability, side reactions and interferences caused by the presence of degradation products and chromatography resolution loss are reduced^{47,48}. The main drawback of using ion-pairing LC-MS is that it may cause ion suppression or contamination of ion source⁴⁹. Taking into account the benefits and handicaps of these analytical methods, the use of ion pairing RPLC-MS was selected to develop a targeted metabolomics method that aims at these amino acids and amino acid derivatives.

4.5. DESCRIPTION OF TARGETED METABOLITES

According to the literature review from section 4.2, 16 metabolites including CNN have been selected for further study. The characteristics of each of the selected analytes are detailed below to understand the background and the biochemical properties of each of them:

✓ ARGININE (ARG):

ARG is a non-essential amino acid that fulfills many functions such as protein biosynthesis, being an intermediate in the urea cycle and being the precursor of some compounds such as nitric oxide, ADMA and SDMA dimethylarginines, CTN, agmatine and other polyamides⁵⁰. In addition, ARG is present in blood, saliva, sweat and urine, as well as in many other organs and tissues including kidney⁵¹.

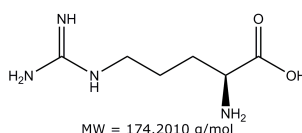


Fig. 4.4. Molecular structure and molecular weight (MW) of ARG.

✓ **ASYMMETRIC DIMETHYLARGININE (ADMA):**

ADMA is derived from ARG and present in blood and urine. ADMA is produced during protein methylation, a post-translational mechanism in which protein modification is performed. More specifically, methyl groups are transferred from some other compounds like SAH⁵², which is also another analyte of interest in this study. ADMA is eliminated in its unaltered form as well as after being metabolized by the action of dimethylarginine dimethylaminohydrolase (DDAH) enzyme in the kidneys⁵³⁻⁵⁶.

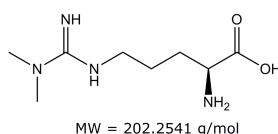


Fig. 4.5. Molecular structure and molecular weight of ADMA.

✓ **BETAINE (BET):**

Betaine, also known as glycine betaine or N,N,N-trimethylglycine, is a small trimethylated amino acid, which is found in blood, urine, breast milk and in several tissues including kidney⁵⁷. BET is present in the organism by the food intake or by oxidation of endogenous choline⁵⁸ and has two key functions in mammals. On the one hand, it works as an osmolyte that is accumulated in almost all of the tissues to help in cell volume regulation and protects the kidneys from possible cellular damage that might occur by hypertonicity^{58,59}. On the other hand, BET is a methyl donor, guaranteeing a proper hepatic function, cellular replication and detoxification reactions^{58,60}.

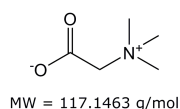


Fig. 4.6. Molecular structure and molecular weight of BET.

✓ **CITRULLINE (CIT):**

CIT is a derivative amino acid, that is, it is produced by metabolic conversions of free amino acids¹³. CIT can be found in plasma, urine, sweat and saliva, as well as in different tissues and organs like the kidney⁶¹. It is originated from ornithine and carbamoyl phosphate, in one of the central reactions of the urea cycle (see Fig. 4.2) as well as from ARG in a different reaction⁶².

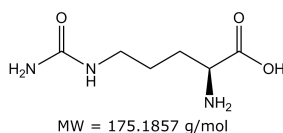


Fig. 4.7. Structure and molecular weight of CIT.

✓ **CREATINE (CTN):**

CTN is a nitrogenous organic acid derived from guanidinoacetate and SAM. In humans CTN can be found in blood, urine, sweat, saliva, breast milk and in many tissues including the kidneys⁶³. It is involved in creatine-creatine phosphate energy transfer as a regulator of ATP/ADP cytosolic ratio in tissues with high or variable use of ATP, such as cardiac and skeletal muscle⁶⁴. CTN is metabolized to CNN, which is excreted by the kidneys⁶⁵.

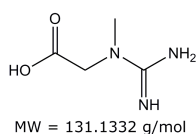


Fig. 4.8. Structure and molecular weight of CTN.

✓ **CREATININE (CNN):**

CNN is the product of decomposition of creatine phosphate in muscle, catalyzed by creatinine amidohydrolase². CNN is usually produced in a constant rate and depends on the muscle mass. It is estimated that 50 mg are generated per kg of muscle daily. Then, this compound is transferred through plasma to the kidneys where it is eliminated by glomerular filtration and partial tubular excretion⁶⁶. CNN can be found in almost all the body fluids, such as blood, urine, sweat, saliva and breast milk, and in many tissues and organs like kidney⁶⁷. If kidneys are impaired, CNN filtering is deficient and thus, serum creatinine levels (SCr) increase^{66,67}.

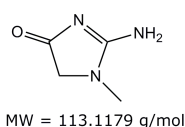


Fig. 4.9. Structure and molecular weight of CNN.

✓ **CYSTEINE (CYS):**

CYS is a sulfur containing amino acid or aminothiols, which can be found in blood, urine, saliva, sweat and some other tissues, such as kidney^{68,69}. Regarding the biochemical properties of CYS, its thiol group undergoes oxidation/reduction reactions, so that if CYS is oxidized cystine originates, a compound formed by two CYS residues gathered together through a disulfide bond. This reaction is reversible and cystine reduction can regenerate two CYS molecules⁶⁹. CYS oxidizes to cystine easily in extracellular oxygenated solution, due to the fact that cysteine sulfhydryl radicals are deprotonated in physiologic pH, thus increasing its reactivity⁷⁰. CYS transport in the organism is carried out mainly by GSH, which will also be studied in current targeted metabolomics method^{71,72}.

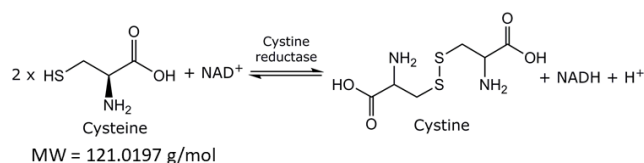


Fig. 4.10. Structure of CYS and cysteine-cystine reaction.

✓ **DIMETHYLGLYCINE (DMG):**

DMG is an amino acid derivative present in blood, urine, saliva and tissues like liver or kidneys⁷³. It is produced from betaine and homocysteine as a consequence of a methyl group transference⁵⁸. Accumulation of DMG, could inhibit the activity of betaine-homocysteine S-methyltransferase enzyme, leading to hyperhomocysteinemia in some cases²⁶. DMG degradation occurs in the mitochondria, being GLY the ultimate metabolite⁵⁸.

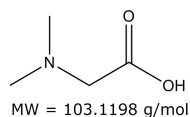


Fig. 4.11. Molecular structure of DMG and its molecular weight.

✓ **GLYCINE (GLY):**

GLY is a non-essential amino acid, located in biofluids like blood, sweat, saliva, urine, bile and feces, in addition to different organs and tissues, such as kidneys⁷⁴. As showed in Figure 4.2, GLY is actively involved in various reactions in the arginine-creatine metabolic pathway, such as the synthesis of GSH or the reaction of GLY and ARG leading to ornithine and guanidinoacetate, catalyzed by AGAT enzyme^{8,68}.

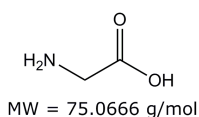


Fig. 4.12. Structure and molecular weight of GLY.

✓ **GLUTATHIONE (GSH):**

Glutathione or γ -glutamyl-cysteinyl-glycine is a water-soluble tripeptide composed of glutamic acid, GLY and CYS, which can be found in blood, urine and saliva⁷⁵. GSH represents the greatest active non-enzymatic intracellular antioxidant defense, acting as a scavenger of reactive oxygen species (ROS), such as hydrogen peroxide and hydroxyl radical, providing them reducing equivalents^{76,77}. GSH might be oxidized combining two reduced GSH molecules by means of a disulfide bond, thus obtaining one oxidized glutathione homodimer (GSSG). This reaction is reversible and GSSG molecule might be reduced to GSH in a reaction mediated by glutathione reductase⁷⁶. In addition, GSH can bind to proteins, resulting in the formation of glutathionylated proteins⁷⁸. A transpeptidase located in the kidneys, γ -glutamyl-transpeptidase (γ -GT), is the major responsible for GSH elimination from plasma⁷⁹.

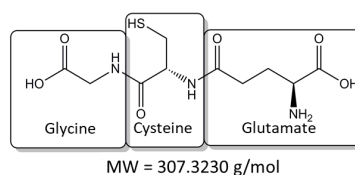


Fig. 4.13. Molecular structure of GSH where the molecular weight of this tripeptide and its three residues are indicated.

✓ **HOMOARGININE (HARG):**

HARG is an amino acid derivative with endogenous origin, that can be found in blood and urine⁸⁰. This compound is excreted through the kidney, where transamination from lysine to HARG occurs⁸¹. It has to be noted that HARG protects against cardiovascular disorders working as a precursor of nitric oxide, a substance related to reduction of endothelial dysfunction⁸²⁻⁸⁴.

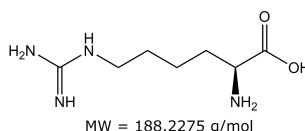


Fig. 4.14. Molecular structure of HARG and its molecular weight.

✓ **HOMOCYSTEINE (HCYS):**

Homocysteine is a sulfur containing amino acid, originated from dietary MET demethylation^{85,86}. This analyte is present in blood, urine and saliva. Besides, multiple tissues like the adipose, brain, fibroblasts, intestine, kidney, liver, muscle tissue, nervous tissue, pancreas and placenta among others also contain homocysteine⁸⁷. Approximately 1 % of homocysteine is estimated to be in plasma in its reduced form (HCYS), whereas the rest is expected in its oxidized form. From oxidized homocysteine, approximately 5 % is present as a homocysteine dimer, homocystine, another 5-10 % as HCYS-CYS dimers and the remaining 80-90 % would be bound to proteins by disulfide bonds⁸⁸. Total fasting plasma levels reflect intracellular metabolism and cellular export of HCYS⁸⁹. Intracellular HCYS can be degraded into CYS, through the transsulfuration metabolic pathway, which takes place in liver and kidney cells⁸⁶.

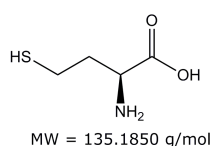


Fig. 4.15. Structure and molecular weight of HCYS.

✓ **METHIONINE (MET):**

MET is an essential amino acid, obtained from diet and required for normal growth and development of the organism. Regarding its location, this analyte can be found in the following biofluids: blood, urine, saliva and sweat⁹⁰. The most important functions of MET include being a substrate for protein synthesis and an intermediate compound in transmethylation reactions, where it works as a methyl donor⁹¹. Finally, MET is metabolized into SAM by methionine adenosyl transferase enzyme (MAT).

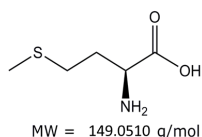


Fig. 4.16. Structure and molecular weight of MET.

✓ **S-ADENOSYLHOMOCYSTEINE (SAH):**

SAH is a potent inhibitor of methylation reactions and could also be indirectly responsible for transmethylation reactions in tissues^{92,93}. SAH may be detected and quantified in blood and urine⁹³. It has to be highlighted that this metabolite is also present in several tissues, such as kidney and liver. Indeed, kidneys play an important role in SAH level regulations,

as they are responsible for its elimination from circulation. SAH is metabolized into HCYS and adenosine⁹⁴.

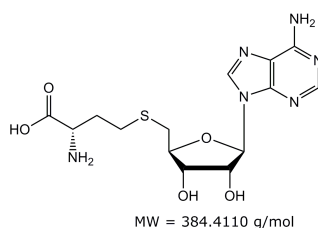


Fig. 4.17. Formula and molecular weight for SAH.

✓ **S-ADENOSYLMETHIONINE (SAM):**

SAM is the product of metabolization of MET and can be found in biofluids like urine and blood, in addition to different organs such as kidneys⁹⁵. SAM is actively involved in transmethylation reactions, where a methyl group is transferred from SAM methyl donor to an acceptor compound and SAH demethylated compound is obtained^{96,97}. SAM/SAH ratio is called methylation index and can be altered by SAM decrease, SAH increase or both^{8,98}.

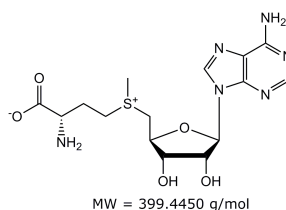


Fig. 4.18. Structure and molecular weight of SAM.

✓ **SYMMETRIC DIMETHYLARGININE (SDMA):**

SDMA is the biologically inactive stereoisomer of ADMA¹⁸, which can be found in blood, urine, saliva and kidneys⁹⁹. SDMA acts as an endogenous inhibitor of arginine-nitric oxide pathway, competing with ARG to introduce into cells⁵⁶. SDMA synthesis is a two-step methylation process starting from ARG, in which SAM acts as a methyl donor¹⁰⁰⁻¹⁰².

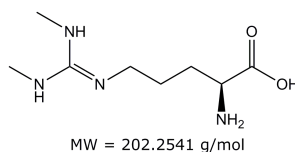


Fig. 4.19. Molecular structure of SDMA and its molecular weight.

4.6. MATERIALS AND EQUIPMENT

4.6.1. STANDARDS AND REAGENTS

All the solvents used for the mobile phase had spectral purity for LC-MS and reagents had analysis quality. For mobile phase and standard preparation acetonitrile from Scharlau (Sentmenat, Spain), methanol from Scharlau (Sentmenat, Spain), ammonium formate from Fluka Analytical, Sigma-Aldrich (Steinheim, Germany), perfluoroheptanoic acid 96 % from Across Organics (New Jersey, USA) and hydrochloric acid from Merck 25 % (Darmstadt, Germany) were acquired. Ultra-high purity water was obtained from tap water pre-treated by Elix reverse osmosis, and subsequent filtration by a Milli-Q system from Millipore (Bedford, MA, USA).

The analytical standards used were obtained from different manufacturers. L-methionine, glycine, L-arginine, L-homocysteine, $N^G, N^{G'}$ -dimethyl-L-arginine di(p-hydroxyazobenzene-*p'*-sulfonate) (SDMA), N^G, N^G -dimethylarginine dihydrochloride (ADMA), S-adenosyl-L-methionine (SAM), S-adenosyl-L-homocysteine (SAH) and citrulline from Sigma-Aldrich (Steinheim, Alemania). Betaine, reduced glutathione, L-cysteine and creatine were supplied by TCI (Tokyo, Japan). On the other hand, N,N-dimethylarginine and homoarginine were purchased from Fluorochem (Hatfield, United Kingdom). Creatinine was obtained from Alfa Aesar (Karlsruhe, Germany) manufacturer. Regarding isotopically labelled compounds creatine- d_3 H₂O was obtained from CDN Isotopes (Quebec, Canada), $N^G, N^{G'}$ -dimethyl-L-arginine- d_6 and creatinine- d_3 were both supplied by Toronto Research Chemicals, TRC-Canada (North York, Canada) and glutathione- $^{13}C_2^{15}N$ obtained from Cambridge Isotope Laboratories (Andover, MA, USA).

In addition, dithiothreitol and 2-mercaptoethanol thiol reducing agents were obtained from Fisher Scientific (Pittsburgh, PA, USA) and Sigma-Aldrich (Steinheim, Germany), respectively. Synthetic Serasub® serum from CST Technologies (Great Neck, NY, USA) was also purchased to be used as a surrogate matrix.

Concerning different solid phase extraction cartridges used, Oasis MCX 3 cc, 60 mg supplied by Waters (Milford, MA, USA) and Hybrid SPE Phospholipid Ultra 30 mg/1 mL from Supelco (Bellefonte, PA, USA) were assayed.

4.6.2. INSTRUMENTAL ANALYSIS

The development of the LC-MS based targeted metabolomics analytical method has been carried out in a HPLC Agilent 1200 Series system coupled to a hybrid quadrupole-time of flight (LC-QTOF) mass spectrometer Agilent 6520 from Agilent Technologies (Santa Clara, CA, USA). This MS detector operated in low mass range (1700 m/z), at 2 GHz extended dynamic range, and centroid mode was used for data collection and storage. LC-QTOF instrument gave a resolution greater than 10000 full widths at half maximum (FWHM) at 118 m/z and greater than 20000 FWHM at 1522 m/z . MS spectra were acquired with a scan rate of 2.2 spectra/s in MS mode and with a scan rate of 3 spectra/s in MS/MS mode.



Fig. 4.20. LC-QTOF equipment used in this study.

Mass accuracy in LC-QTOF is guaranteed by means of direct infusion into the source of purine (121.0509 m/z) and HP-921 (922.0098 m/z), enabling continuous internal calibration during analysis and ensuring accuracy and reproducibility.

During method optimization 3 different chromatographic columns were assayed: Gemini-NX 110A C18 (4.6 x 150 mm, 5 μm) from Phenomenex (Torrance, CA, USA), Poroshell 120 EC-C18 (4.6 x 100 mm, 2.7 μm) from Agilent (PA, USA) and Zorbax Eclipse plus C8 (2.1 x 150 mm, 5 μm) from Agilent (PA, USA).

Mobile phases were filtered through de Millipore Omnipore (Watford, Ireland) 0.1 μm filters and degassed with a Digital Ultrasonic Cleaner CD-4820 (Medellín, Colombia) ultrasonic system. The pH from the mobile phase was measured with a GLP 21 Crison pH-meter (Alella, Spain).

4.6.3. PREPARATION OF STANDARD SOLUTIONS

First of all, stock standard solutions containing around 1000 mg/L were prepared and from these solutions individual standard solutions of 1 mg/L, 2 mg/L and 5 mg/L were made for precursor ion identification. Stock solutions of MET, GLY, ARG, CYS and GSH stock standard solutions were dissolved in water. In order to contribute to their stability, BET, CTN, CIT, DMG, HARG, CNN, SAH, HCYS, SAM and ADMA were prepared in 0.01 N of HCl. Finally, SDMA was dissolved in methanol. Concerning internal standards, creatinine-d₃ and SDMA-d₆ were prepared in methanol, whereas creatine-d₃ and glutathione-¹³C₂¹⁵N were dissolved in water. These solutions were stored frozen at -40°C.

Intermediate mixed solutions were prepared weekly in ammonium formate 5 mM and stored frozen in aliquots. Working solutions were obtained from intermediate mixed solutions for analytical methodology and sample treatment optimization. Due to the observed degradation of working solutions at room temperature, there was a need for replacing them daily with frozen intermediate solution.

Calibration standards were prepared in pooled plasma made of CKD and control samples, which had been previously quantified by standard addition and contained: 159 ng/mL ADMA, 14.3 µg/mL ARG, 3.1 µg/mL BET, 2.8 µg/mL CIT, 9.7 µg/mL CTN, 15.2 µg/mL CNN, 13.6 µg/mL CYS, 1.0 µg/mL DMG, 3.2 µg/mL GSH, 1.8 µg/mL GLY, 0.4 µg/mL HARG, 0.5 µg/mL HCYS, 2.7 µg/mL MET, 36 ng/mL SAH, 55 ng/mL SAM and 291 ng/mL SDMA. Pooled plasma was doped to yield concentrations at different ranges depending on the compound (see section 4.8.1).

Similarly, quality control samples (QC) consisted in a pooled plasma quantified beforehand, unspiked and spiked at low, medium and high levels. Unspiked plasma contained: 95 ng/mL ADMA, 10.6 µg/mL ARG, 1.2 µg/mL BET, 1.7 µg/mL CIT, 4.4 µg/mL CTN, 5.2 µg/mL CNN, 7.9 µg/mL CYS, 0.8 µg/mL DMG, 1.6 µg/mL GSH, 1.3 µg/mL GLY, 0.6 µg/mL HARG, 0.4 µg/mL HCYS, 1.3 µg/mL MET, 17 ng/mL SAH, 16 ng/mL SAM and 98 ng/mL SDMA. High QC consisted in the addition of 12.5 µg/mL of GSH, GLY, DMG, BET, CTN, MET and CNN, 75 µg/mL of CYS, 25 µg/mL of ARG and CIT, 2.5 µg/mL of HARG and HCYS, 0.4 µg/mL of ADMA and SDMA, and 0.25 µg/mL of SAH and SAM. Regarding low and medium QC, same pooled plasma was used but spiked with the 10 % and the 50 % of the concentration added for each analyte in high QC level.

4.6.4. SUBJECTS AND SAMPLING

Thirty-two patients suffering from CKD, aged 3-18 years, and twenty-four control patients, aged 6-19 years were recruited for the study. Paediatrics suffering from CKD were at different degrees of the disease, as showed in Table 4.2. All paediatric CKD patients were followed up in Cruces University Hospital (Bilbao, Spain) and were clinically stable at the time of study. The exclusion criteria for this study included patients with anuria, hepatopathy and/or insulin-dependent diabetes mellitus. Regarding control patients, they met the following inclusion criteria: healthy children who had a minor surgery in Cruces University Hospital.

Table 4.2. Characteristics of the population subjected to study.

Characteristics of the population				Number of patients
CKD DEGREE	SEX (M/F)	AGE (2-12 y/13-18 y)	TREATMENT (No RRT/ Dialyzed/ Transplanted)	
CONTROL	18/6	15/9	24/0/0	24
CKD2	8/6	10/4	9/0/5	14
CKD3	5/1	2/4	4/0/2	6
CKD4	2/4	3/3	5/0/1	6
CKD5	2/4	5/1	1/5/0	6

Blood samples were collected in the morning after an overnight fasting. Samples were cooled immediately after blood collection in an ice-water bath and centrifuged at 1000 g for 5 min at 4°C. Finally, samples were stored at -80°C until sample treatment and analysis.

Ethics Committee of Clinic Research of Cruces Hospital approved the study protocol and patients' parents gave written informed consent.

4.7 OPTIMIZATION OF THE ANALYTICAL METHOD

In general, amino acids, amino acid derivatives and related compounds mentioned above are characterized for being small compounds with polar nature, even though there are some exceptions in size, e.g. amino acids bounded to nucleosides.

Traditionally, the most popular techniques for the analysis of amino acids and amino acid derivatives have been the following ones:

- ✓ Cation-exchange chromatography with ultraviolet detector (IEC-UV), recently replaced by automatic analyzers
- ✓ Reverse-phase liquid chromatography coupled to fluorescence (LC-FLD), ultraviolet (LC-UV), diode array (LC-DAD) or mass spectrometry detectors (LC-MS)
- ✓ Hydrophilic interaction liquid chromatography coupled to mass spectrometry (HILIC-MS)
- ✓ Ion-pairing-reverse phase liquid chromatography coupled to mass spectrometry (IP-LC-MS)
- ✓ Gas chromatography coupled to mass spectrometry (GC-MS)

The most relevant research studies published over the last 15 years about amino acid and amino acid derivative determination are gathered in Table 4.3.

According to literature, a decade ago GC-MS using different derivatizers¹⁰³⁻¹⁰⁶, ion-pairing LC-MS techniques^{47,107,108} and automatic analyzers based on classical cation-exchange chromatography with ultraviolet detection^{109,110} were the most frequently used analytical techniques for amino acid determination. However, recent research studies show that trend during the last years goes towards using LC-MS techniques, either HILIC¹¹¹⁻¹¹⁴ or ion-pairing RPLC^{115,116}.

Table 4.3. Literature review on analytical methods used for the analysis of amino acids, amino acid derivatives and related compounds.

Analytes	Reagent	Analytical technique	Author	Year
CYS and GSH	-	HILIC- LC-ESI-QQQ	Cao ¹¹¹	2015
CNN	Ethylchloroformate	RP-LC-UV	Leung ¹¹⁷	2014
CIT and ARG	PFHA	UHPLC-ESI-LTQ	Oosterink ¹¹⁵	2014
HCYS, SAH and SAM	-	HILIC-LC-Q-Orbitrap	Ngo ¹¹²	2014
ARG	MSTFA/MBTFA	GC-MS	Yoon ¹¹⁸	2013
ARG, ADMA, SDMA, HARG	-	HILIC-LC-ESI-QQQ	Martens-Lobenhoffer ^{113,119,120}	2013, 2012, 2007
GLY	-	HILIC-LC-ESI-QQQ	Tang ¹²¹	2012
ADMA and SDMA	PFPA and methanol	GC-MS	Tsikis ¹²²	2011
HARG	1-butanol	RP-LC-MS	Atzler, ¹²³	2011
CTN, CNN, CIT, BET	-	HILIC-UPLC-TOF-MS	Spagou ¹¹⁴	2011
ARG, ADMA, SDMA and CIT	-	HILIC-LC-ESI-QQQ	Brown ¹²⁴	2011
ARG, CYS, HCYS, CNN, MET, GLY and CIT	Ninhydrin, OPA, MTBSTFA	Biochrom 30 (IEC-UV), RP-LC-MS and GC-MS	Andrade ⁸	2011
ARG, CIT, GLY, MET and CYS	HFBA	UHPLC-APCI-QQQ	Harder ¹¹⁶	2011
BET, DMG	-	HILIC-UPLC-ESI-QQQ	Kirsch ¹²⁵	2010
ARG, GLY and MET	FMOC	RP-LC-DAD/FLD	Jámbor ¹²⁶	2009
ARG	Ninhydrin	Biochrom 30 (IEC-UV)	Loscos ¹¹⁰	2008
ARG, CYS, CIT, GLY, HCYS and MET	Ninhydrin	Beckman 7300 (IEC-UV)	Dietzen ¹⁰⁹	2008
GLY and MET	HFBCF y PFPCF	GC-MS	Husek ¹⁰⁶	2008
SAM	-	HILIC-LC-QTRAP	Wang ¹²⁷	2008
GLY, ARG and MET	OPA	RP-LC-FLD	Yokoyama ¹²⁸	2008
ADMA	PFPA	GC-MS	Beckmann ¹⁰⁵	2008
CIT	HFBA and propanol	GC-MS	Rougé ¹⁰⁴	2008
GLY and MET	MSTFA	GC-MS	Kaspar ¹⁰³	2008
CTN and CNN	-	HILIC-LC-UV/ ESI-MicroTOF	Godejohann ¹²⁹	2007
GLY, CYS, MET, ARG and CIT	PFHA	LC-ESI-TOF	Armstrong ¹⁰⁷	2007
GLY, DMG, CYS, HCYS, MET, ARG, CIT, GSH, SAM and SAH	PFHA	LC-ESI-QQQ	Piraud ⁴⁷	2003
GLY, CYS, MET, ARG and CIT	PFHA and TCA	LC-turboionspray (TIS)-QQQ	Qu ¹⁰⁸	2002

PFHA = perfluoroheptanoic acid, MSTFA = N-Trimethylsilyl-N-methyl trifluoroacetamide, MBTFA = N-Methyl-bis(trifluoroacetamide), OPA = o-phthalaldehyde, HFBA = Heptafluorobutyric acid, MTBSTFA = N-Methyl-N-tert-butyl dimethylsilyltrifluoroacetamide, HFBCF = Heptafluorobutyl chloroformate, PFPCF = 2,2,3,3,3-pentafluoropropyl chloroformate, FMOC = Fluorenylmethyloxycarbonyl

Taking into account the advantages and inconveniences of each of these techniques (described in section 4.4) in addition to the trend towards using LC-MS techniques, ion pairing RPLC-MS has been selected to develop a new targeted metabolomics method by means of LC-QTOF equipment. This development includes the optimization of the chromatographic variables, selection of mass spectrometry variables as well as the optimization of sample treatment.

4.7.1. SELECTION OF THE CHROMATOGRAPHIC VARIABLES

The first step consisted in the analysis and characterization of the sixteen analytes under study, one by one, obtaining their retention times and their characteristic m/z . The preliminary assays were carried out using 5 mM ammonium formate and 0.5 mM PFHA in aqueous phase (A) and acetonitrile in organic phase (B) for chromatographic separation. In addition, ESI interface in positive ion mode and the following mass spectrometry conditions were set: capillary voltage, 3000 V; drying gas temperature, 325 °C; drying gas flow 10 L/min; nebulizer pressure, 40 psig; and fragmentor voltage, 125 V. In all cases, the protonated form of each metabolite, $[M+H]^+$, was obtained and exact masses for the analytes of interest are showed in Table 4.4.

Table 4.4. Molecular formulas, molecular exact masses, ionization and accurate m/z ratios in MS for analytes of interest.

Analyte	Molecular formula	MW	Ionization	m/z
ADMA	$C_8H_{18}O_2N_4$	202.1430	$[M+H]^+$	203.1503
ARG	$C_6H_{14}O_2N_4$	174.1117	$[M+H]^+$	175.1193
BET	$C_5H_{11}O_2N$	117.0790	$[M+H]^+$	118.0865
CIT	$C_6H_{13}O_3N_3$	175.0957	$[M+H]^+$	176.1030
CTN	$C_4H_9O_2N_3$	131.0695	$[M+H]^+$	132.0768
CNN	$C_4H_7ON_3$	113.0589	$[M+H]^+$	114.0667
CYS	$C_3H_7O_2NS$	121.0197	$[M+H]^+$	122.0270
DMG	$C_4H_9O_2N$	103.0633	$[M+H]^+$	104.0706
GSH	$C_{10}H_{17}O_6N_3S$	307.0838	$[M+H]^+$	308.0911
GLY	$C_2H_5O_2N$	75.0320	$[M+H]^+$	76.0393
HARG	$C_7H_{16}O_2N_4$	188.1273	$[M+H]^+$	189.1346
HCYS	$C_4H_9O_2NS$	135.0354	$[M+H]^+$	136.0427
MET	$C_5H_{11}O_2NS$	149.0510	$[M+H]^+$	150.0583
SAH	$C_{14}H_{20}O_5N_6S$	384.1216	$[M+H]^+$	385.1289
SAM	$C_{15}H_{22}O_5N_6S$	398.1372	$[M+H]^+$	399.1445
SDMA	$C_8H_{18}O_2N_4$	202.1430	$[M+H]^+$	203.1503

4.7.1.1. Column

Regarding that these analytes are characterized for being in general small polar compounds and some of them hardly retain, according to preliminary experiments carried out, three different chromatographic columns were assayed: Gemini-NX 110Å C18 (4.6 x 150 mm, 5 µm) from Phenomenex (Torrance, CA, USA), Poroshell 120 EC-C18 (4.6 x 100 mm, 2.7 µm) and Zorbax Eclipse Plus C8 (2.1 x 150 mm, 5 µm). First step consisted in analyzing individually all the analytes using these columns to get their retention times.

First of all, Gemini-NX 110Å C18 was experimented and a number of different mobile phase gradient programs were studied. However, as a consequence of the similarity of some of the analytes of interest, multiple amino acids coeluted. In addition, peak resolution was not good and wide peaks were obtained. For that reason, different column temperatures were assayed to check whether peak shape could be improved using higher temperatures: 30°C, 35°C and 40°C. Nevertheless, due to the impossibility to enhance the resolution of the peaks to an acceptable level, testing a different chromatographic column was decided.

Zorbax Eclipse Plus C8 (2.1 x 150 mm, 5 µm) was assayed next. Results for this column were better than for Gemini-NX as it allowed a better separation. However, peak resolution was still insufficient and therefore, experimenting with a different column was decided.

Finally, Poroshell 120 EC-C18 column (4.6 x 100 mm, 2.7 µm) was tested. From the first moment, this column thermostated at 30 °C proportioned a higher resolution and at the same time improved the retention of different compounds taking into account their similarity and their trend to coelute or to elute at similar retention times. For all these reasons, Poroshell 120 EC-C18 column was selected and different gradients and flow rates were assayed, as detailed in the following section.

4.7.1.2. Composition of mobile phase

Due to the fact that amino acids and amino acid derivatives are polar compounds that in general hardly retain, the use of ion-pairing liquid chromatography was selected for separation purposes. Ion pairing reagents are large ionic molecules containing an alkyl chain, which provide certain hydrophobicity allowing the retention of the ion pair on a reversed-phase column. The addition of ion pairing reagents to liquid chromatography mobile phases in reverse-phase mode can increase selectively the retention of charged analytes¹³⁰.

Perfluorocarboxylic acids are the most frequent ion-pairing reagents used to enhance the retention as they are compatible with electrospray ion source in mass spectrometry¹³¹. Different perfluorinated carboxylic acids have been used in literature for amino acid or related compounds analysis like trifluoroacetic acid (TFA, C₂), heptafluorobutyric acid (HFBA, C₄), nonafluoropentanoic acid (NFPA, C₅), perfluoroheptanoic acid (PFHA, C₇) and pentadecafluorooctanoic acid (PDFOA, C₈)¹³².

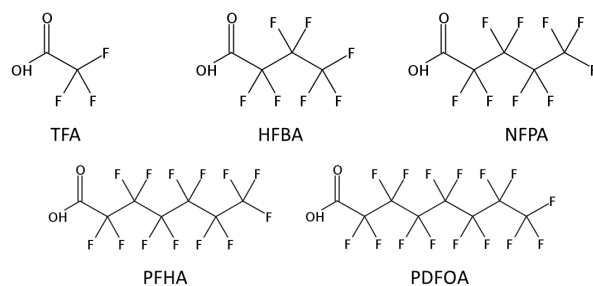


Fig. 4.21. Examples of perfluorinated ion-pairing agents.

In this case, perfluoroheptanoic acid (PFHA) ion pairing reagent was added at a concentration of 0.5 mM in the aqueous phase. This concentration was chosen according to literature in which 0.5 mM was used for different amino acid related applications^{49,107,133}.

Regarding that the analytes of interest are very polar compounds, ESI was selected. It has to be noted that when using ESI interface, appropriate pH adjustment is critical in order that compounds are ionized in the mobile phase. At neutral pH, amino acids behave as "zwitterion" dipoles with charged carboxylate and amino groups. The dissociation constant for the charged carboxylate is $pK_1 = 1.82$ to 2.35 depending on the amino acid, and for the amino groups is $pK_2 = 8.70$ to 10.70 . Moreover, some amino acids, like ARG, contain charged groups in the side chains R^{13} . Taking this into consideration, ammonium formate 5 mM was chosen to be added to the aqueous phase, thus obtaining pH 4.2, which enhances the ionization of most of these compounds.

On the other hand, the possibility of adding PFHA ion-pairing reagent to the organic phase was considered as well, in order to avoid retention time shift within a batch. However, this idea was discarded, since once conditioned the column doing some previous injections, the retention times remain stable regardless using or not using the ion-pairing reagent also in the organic phase. In addition, ammonium formate was not included in the organic phase, as ammonium formate is sparingly soluble in acetonitrile.

Therefore, a mobile phase consisting of ammonium formate 5 mM and 0.5 mM PFHA in the aqueous phase (A) and acetonitrile in the organic phase (B) was selected.

Taking into account the high polarity and the similar physicochemical properties of the majority of the analytes, there was a need for the optimization of a gradient program for chromatographic separation. A large number of gradients were assayed modifying the organic phase percentage. The effect of the flow was also analyzed by comparing 3 different gradients: 0.3, 0.4 and 0.5 mL/min (Fig. 4.22). These flows are within the range

of flows recommended by the commercial brand of the equipment for ESI ionization interface.

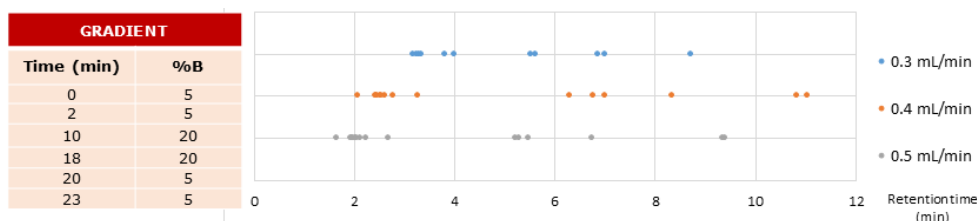


Fig. 4.22. Separation of the analytes according to the flow used for the same gradient.

Despite the fact that in reverse-phase liquid chromatography lower flows generally show the better separations and higher flows better resolution, in this case, similar resolution of the peaks was obtained for all the flows assayed and the best separation was achieved with flows of 0.4 mL/min and 0.5 mL/min (Fig. 4.22). This could be in relation with the use of an ion-pairing reagent. Finally, 0.4 mL/min flow was selected, as proportioned the best separation and, in addition, constitutes an intermediate flow that may favour better ionization in ESI and consumes less solvent.

After assaying a number of gradients, the chromatographic conditions and the gradient program showed in Table 4.5 allowed to get the best separation and resolution.

Table 4.5. Chromatographic conditions selected for this method.

Chromatographic conditions	
Column	Poroshell 120 EC-C18 (4.6 x 100 mm, 2.7 µm) Poroshell LC inline filter (4.6 mm, 0.2 µm)
Column temperature	30 ± 0.8 °C
Mobile phase	A: 0.5 mM PFHA and 5 mM ammonium formate B: acetonitrile
Flow	0.4 mL/min
Gradient program	0 min, 5 %B 2 min, 5 %B 10 min, 20 %B 11 min, 60 %B 13 min, 60 %B 15 min, 5 %B 17 min, 5 %B Post-time: 2 min
Injection volume	2 µL, needle wash with isopropanol:water (50:50, v/v)

Even though a high number of gradients were experimented, all of them showed some defined compounds that were retained around minute 2.5 in the chromatogram: CYS, DMG, BET, CTN, HCYS, GSH and GLY. It has to be noted that despite this not being the ideal, it is not a problem at all when working with mass spectrometry, as each analyte has

a defined m/z . The problem comes up when isomers with same m/z need to be measured. This is the case of the isomers ADMA and SDMA, both ionizing in their protonated form and leading to the same ion: 203.1503 m/z . None of the gradients tested enabled a proper separation of these isomers. However, different authors pointed out that despite ADMA and SDMA having identical ionization in MS mode and eluting at the same retention time, their MS/MS fragmentation provides unique and specific ion products^{119,134,135}. Therefore, it was decided that these two coeluting analytes would be analyzed and quantified in MS/MS mode.

A chromatogram obtained from the analysis of the analytes of interest with the selected conditions in MS mode is showed in Fig. 4.23.

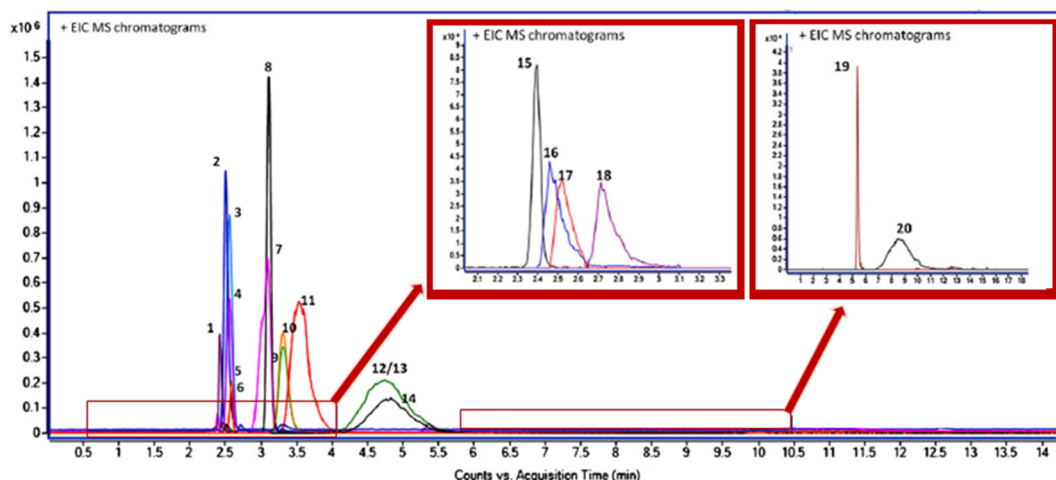


Fig. 4.23. EIC chromatogram corresponding to analytical standard solution in MS mode: 1. CIT 1.5 $\mu\text{g/mL}$; 2. BET 0.75 $\mu\text{g/mL}$; 3. CTN 1.5 $\mu\text{g/mL}$; 4: CTN- d_3 0.75 $\mu\text{g/mL}$; 5. GSH 9 $\mu\text{g/mL}$; 6. GSH- $^{13}\text{C}_2^{15}\text{N}$ 7.5 $\mu\text{g/mL}$; 7. ARG 1.5 $\mu\text{g/mL}$; 8. MET 9 $\mu\text{g/mL}$; 9. CNN 0.75 $\mu\text{g/mL}$; 10. CNN- d_3 0.75 $\mu\text{g/mL}$; 11: HARG 3 $\mu\text{g/mL}$; 12. ADMA 0.2 $\mu\text{g/mL}$; 13. SDMA 0.2 $\mu\text{g/mL}$; 14. SDMA- d_6 0.2 $\mu\text{g/mL}$; 15. GLY 9 $\mu\text{g/mL}$; 16. DMG 9 $\mu\text{g/mL}$; 17. CYS 12 $\mu\text{g/mL}$; 18. HCYS 3 $\mu\text{g/mL}$; 19. SAH 0.2 $\mu\text{g/mL}$; 20. SAM 0.2 $\mu\text{g/mL}$.

4.7.2. SELECTION OF THE MASS SPECTROMETRY VARIABLES.

Once the separation of the analytes was assayed, different variables from the mass spectrometer were optimized for MS spectra and MS/MS spectra analysis.

4.7.2.1. MS spectra

First of all, different interfaces were experimented with the selected chromatographic conditions from Table 4.5. A comparison between Agilent Jet Stream® ESI (AJS ESI) and conventional ESI was carried out using the following preliminary MS conditions for both interfaces: capillary voltage, 3000 V; fragmentor voltage, 125 V; nebulizer pressure, 40 psig; drying gas temperature, 300°C; drying gas flow, 10 L/min; mass range, 50-1000 m/z; MS scan rate, 2.2 spectra/s.

AJS ESI is based on thermal gradient technology, which theoretically enhances the signal of analytes with the help of an auxiliary nitrogen gas with concentric orientation towards the nebulizer spray (sheath gas), reducing ion dispersion at normal flow rate and decreasing the amount of neutral solvent clusters. For AJS ESI interface, sheath gas temperature of 325 °C, a sheath gas flow of 11 L/min and a nozzle voltage of 0 V were also set. Fig. 4.24 shows the chromatograms resulting from the analysis of the standard solutions in AJS ESI and ESI interfaces.

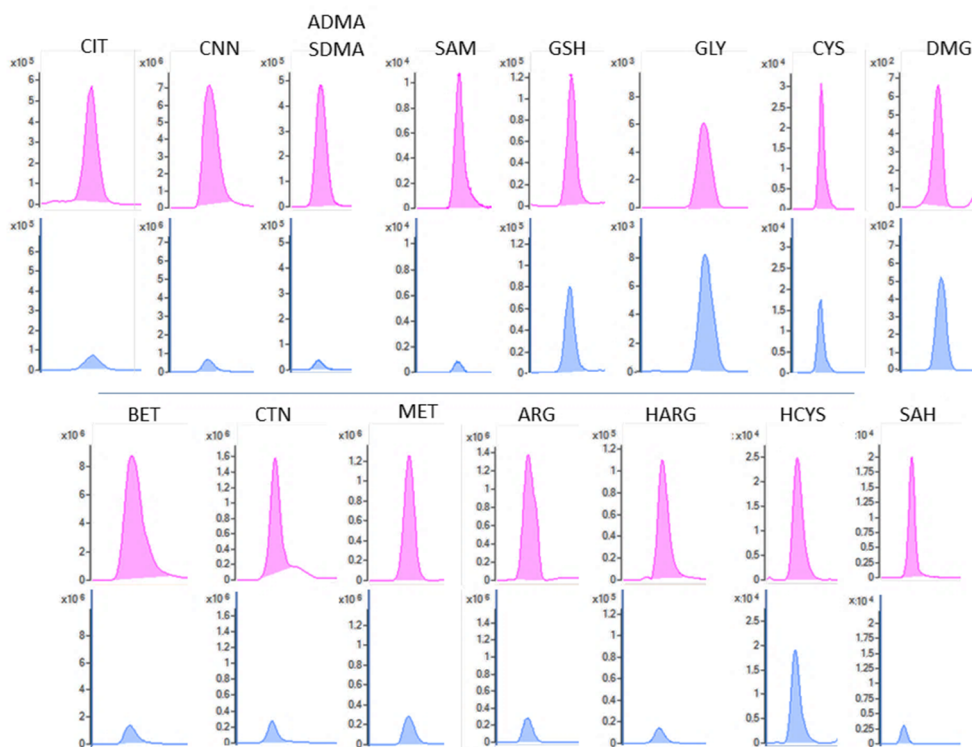


Fig 4.24. Comparison of the signals obtained for each analyte using AJS ESI (pink) or conventional ESI (blue).

All the metabolites except for GLY have higher signals using AJS ESI in comparison with conventional ESI, and regarding GLY, the difference was not significant. Therefore, using AJS ESI interface was selected.

Then, working conditions were optimized for AJS ESI to maximize the intensity of the signals. Sheath gas temperatures of 300 °C, 325 °C and 350 °C were studied, as well as sheath gas flows of 10 L/min and 11 L/min. Similarly, drying gas temperatures of 300 °C, 325 °C and 350 °C and flows of 10 L/min and 11 L/min were also tested. In addition, the effect of nozzle voltages of 0 V, 500 V and 1000 V in analyte signals was studied. Reaching a compromise, the following conditions were set as they favour less sensitive analytes: dry gas temperature, 300°C; dry gas temperature, 10 L/min; sheath gas temperature, 325 °C; and sheath gas flow, 11 L/min.

In addition, fragmentor voltages were optimized, being tested voltages ranging from 75 V to 125 V. Finally, a fragmentor voltage of 75 V was selected, as it was adequate for all the compounds and increased especially the signal of DMG, which decreased substantially using stock solutions with voltages higher than 100 V (see Fig. 4.25) and would be hardly detected in plasma.

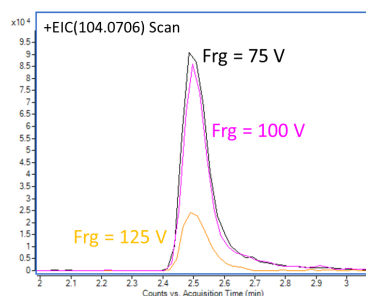
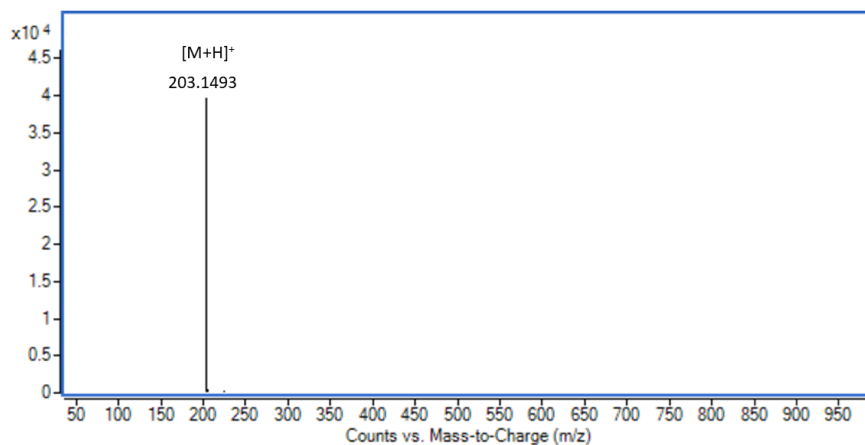


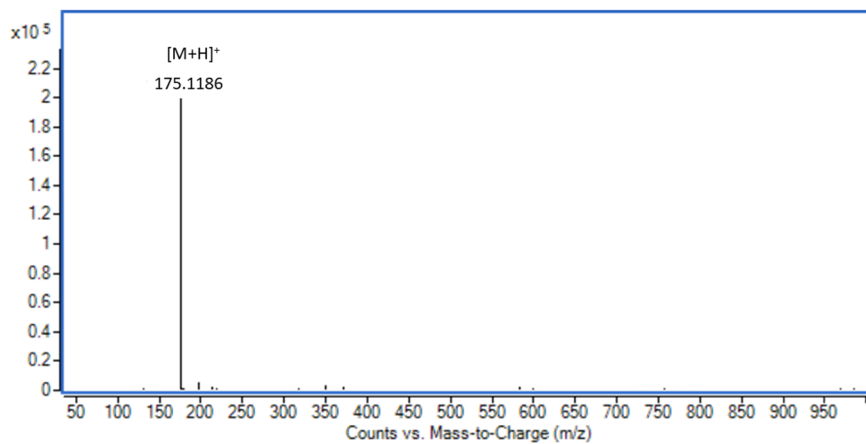
Fig 4.25. Comparison of EIC of dimethylglycine with different fragmentor voltages.

These MS conditions resulted in the following MS spectra for stock solutions containing 5 and 10 $\mu\text{g/mL}$ of the analytes:

✓ **ADMA:**



✓ **ARG:**



✓ **BET:**

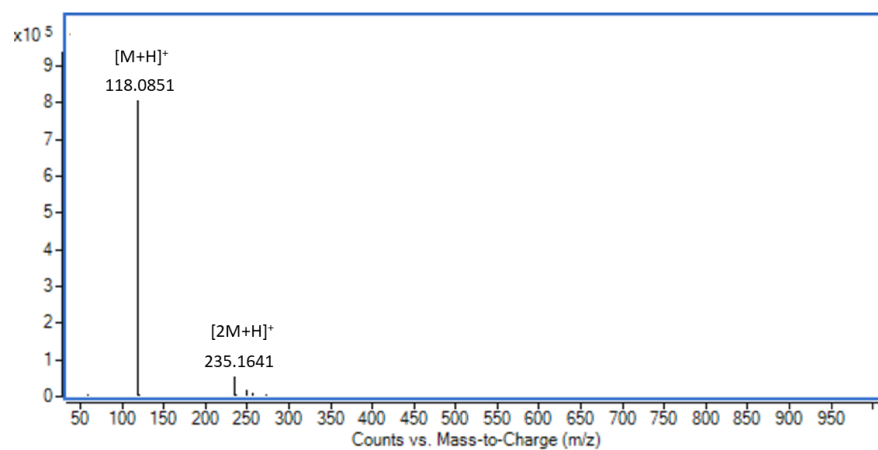
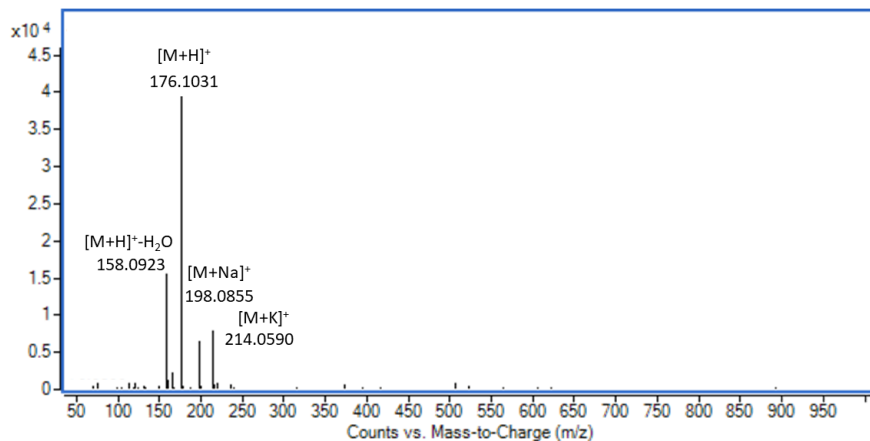
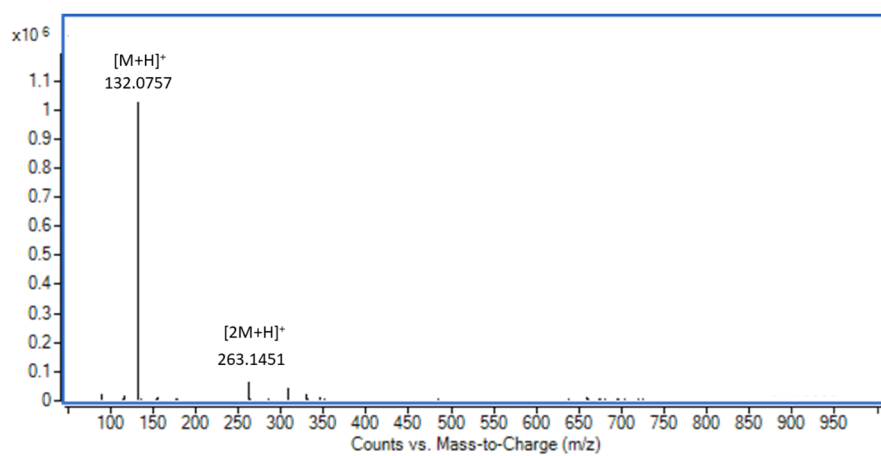
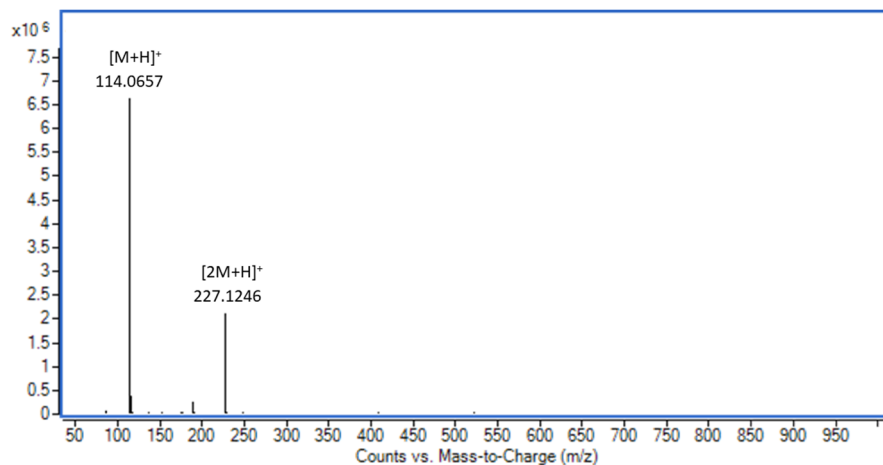
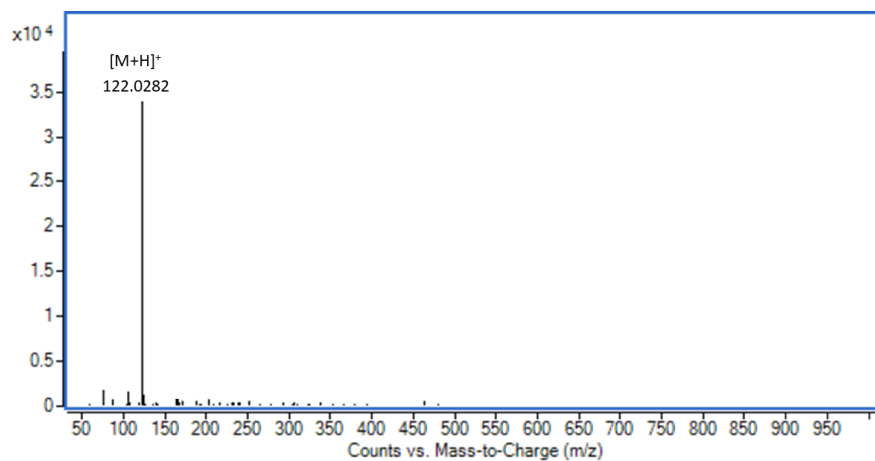


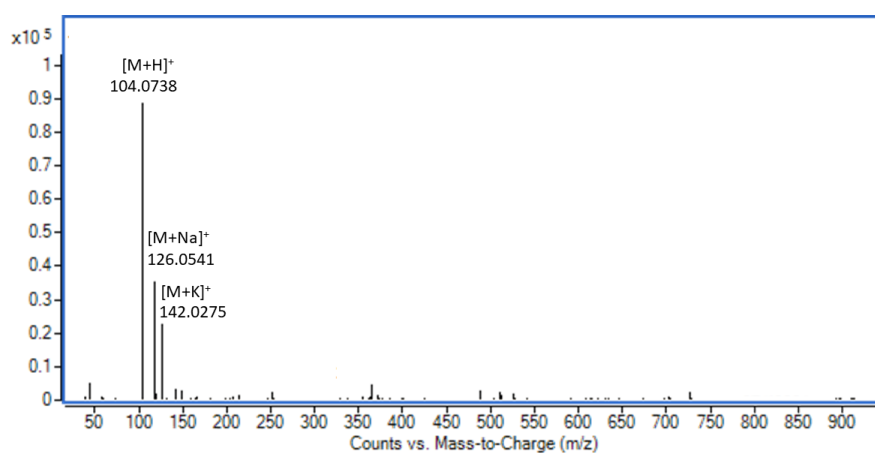
Fig. 4.26. MS spectra of the analytes of interest.

✓ **CIT:**✓ **CTN:**✓ **CNN:****Fig. 4.26 (cont.).** MS spectra of the analytes of interest.

✓ **CYS:**



✓ **DMG:**



✓ **GSH:**

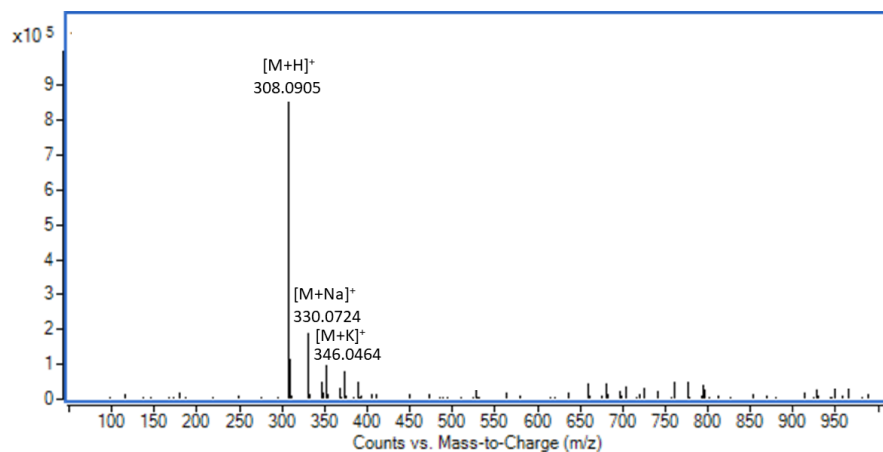
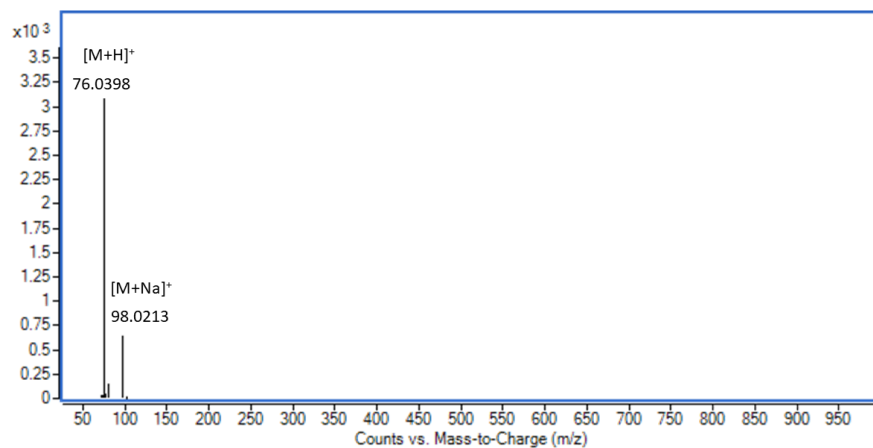
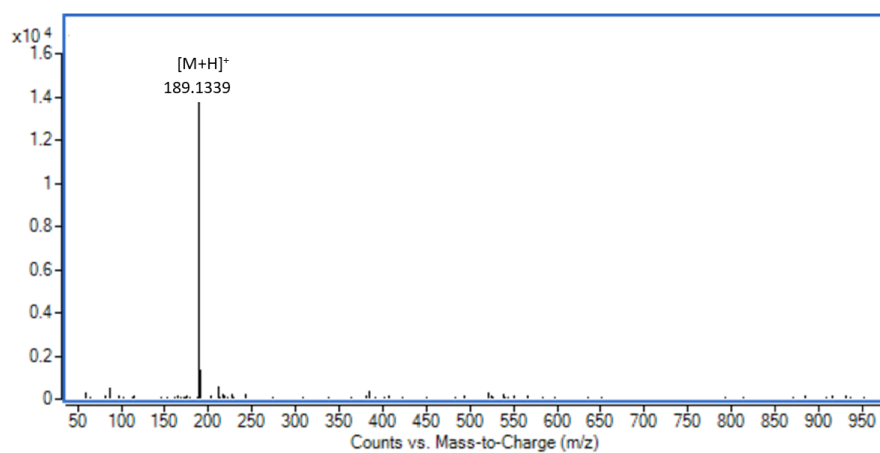
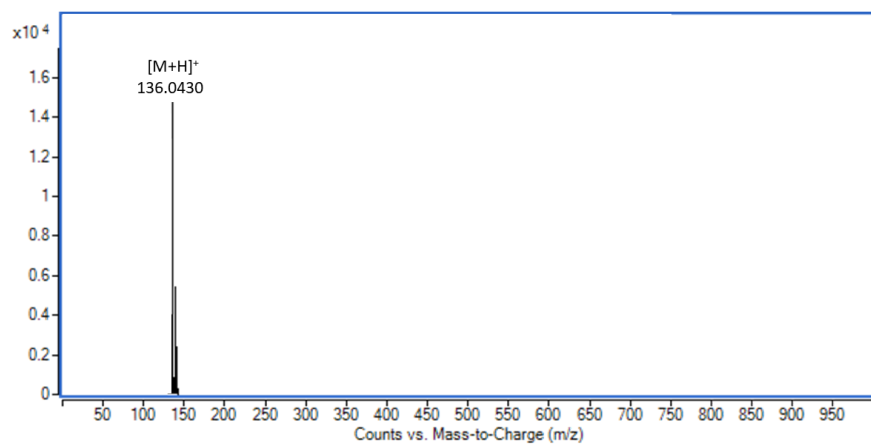
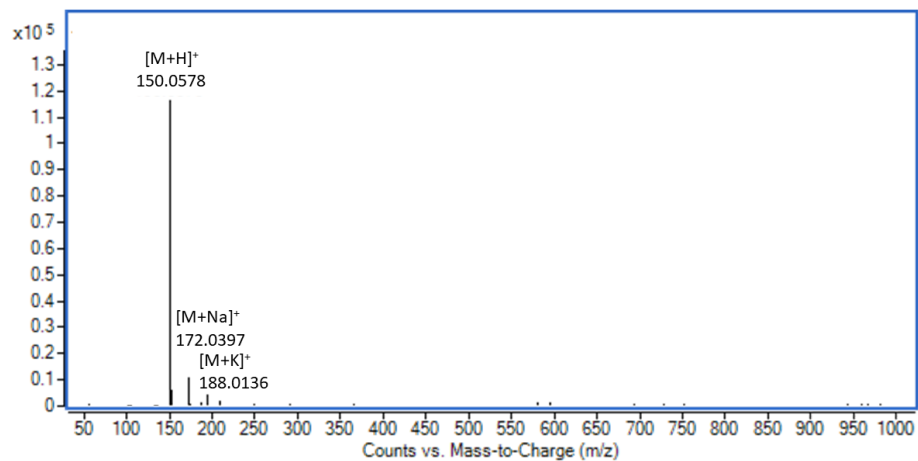


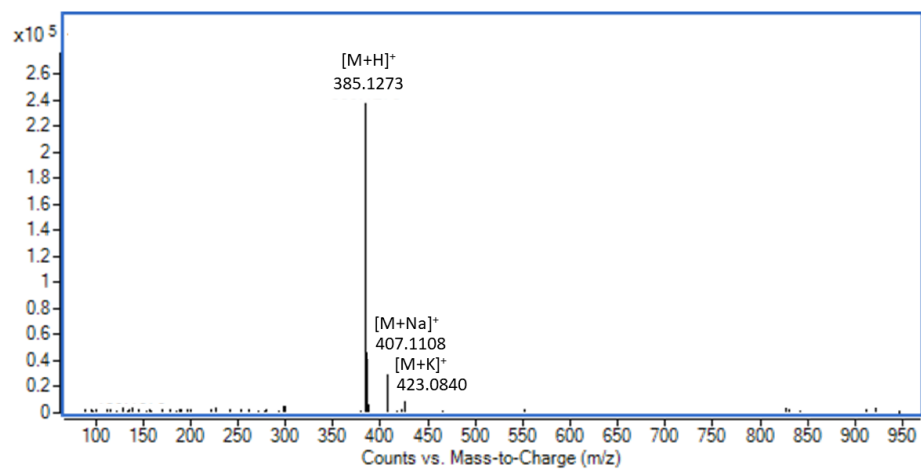
Fig. 4.26 (cont.). MS spectra of the analytes of interest.

✓ **GLY:**✓ **HARG:**✓ **HCYS:****Fig. 4.26 (cont.).** MS spectra of the analytes of interest.

✓ **MET:**



✓ **SAH:**



✓ **SAM:**

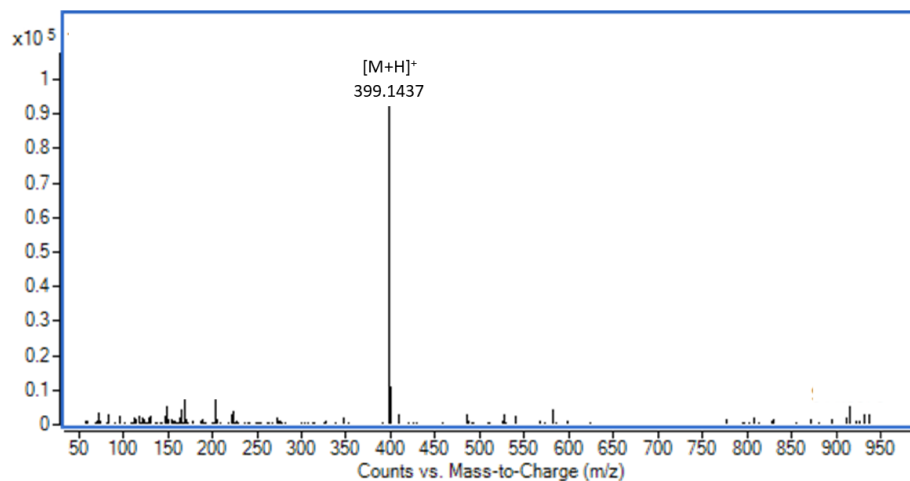


Fig. 4.26 (cont.). MS spectra of the analytes of interest.

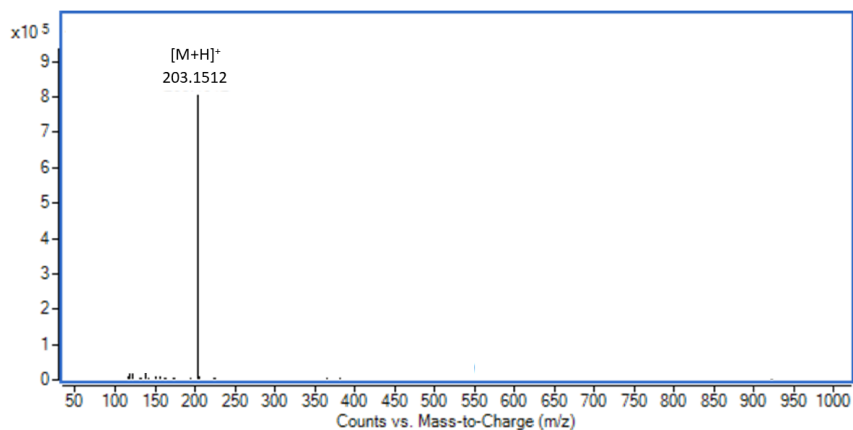
✓ **SDMA:**

Fig. 4.26 (cont.). MS spectra of the analytes of interest.

4.7.2.2. MS/MS spectra

As it can be observed from MS spectra in Fig. 4.26, ADMA and SDMA have the same MS spectra, in addition to eluting at the same time. Therefore, there is a need for MS/MS spectra fragmentation, to obtain unique and specific ion products for ADMA and SDMA, as explained in section 4.7.1.2.

The same conditions from MS mode were used, in 30-1000 m/z mass range and aimed at distinguishing ADMA and SDMA, ramping collision energies from 10 V to 35 V were assayed to obtain characteristic and abundant product ions in MS/MS spectra. The monitorization of unique and specific transitions for ADMA and SDMA was achieved with a collision energy of 10 V: 203.1503>46.0657 for ADMA and 203.1503>172.1086 for SDMA (see Fig. 4.27).

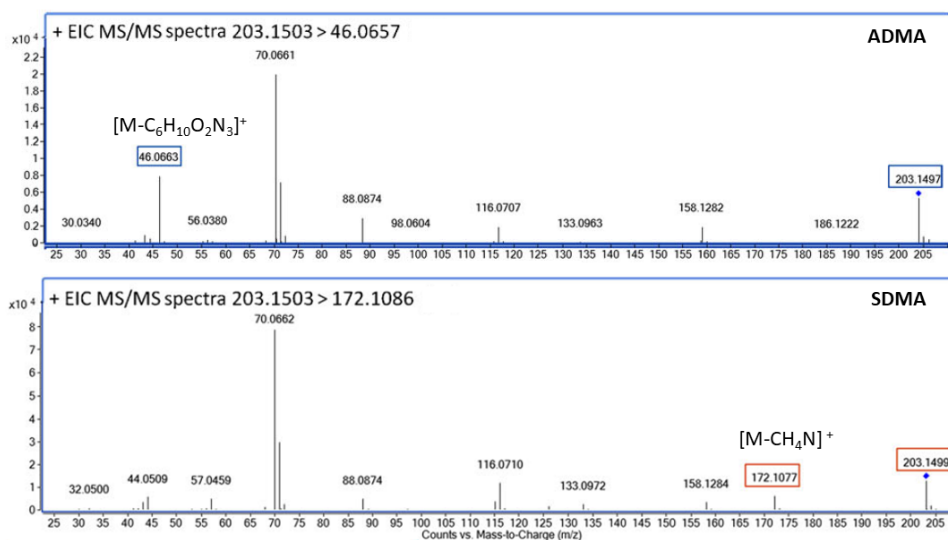


Fig. 4.27. EIC MS/MS spectra showing characteristic transitions for ADMA and SDMA.

In addition, regarding that this instrument when operating in targeted MS/MS mode also analyzes sequentially in MS mode, the number of acquired spectra is reduced in each mode. Therefore, for quantification purpose it is necessary to reach a compromise between the number of acquired spectra and the sensitivity obtained by adjusting the acquisition rate. For that reason, regarding that MS/MS signals are less abundant than MS signals and contain limited number of acquired spectra, MS/MS acquisition rates of 2 spectra/s, 3 spectra/s and 4 spectra/s were assayed to improve the relation between sensitivity and number of acquired spectra. Finally, using 3 spectra/s resulted optimal for MS/MS acquisition and was selected.

To sum up, all the metabolites were analyzed in MS mode, except for ADMA and SDMA, which were acquired in MS/MS mode. Mass spectrometry conditions selected are summarized in Table 4.6. The retention time of each analyte in plasma and accurate mass measurement of the precursor ion with an error less than 5 ppm (MS spectra) or accurate mass measurement of the selected product ion (MS/MS spectra) for the analytes measured in MS/MS mode were the necessary mainstays for the positive identification of the target compounds. According to the EU's decision 2002/657/CE proposed for the collection of identification points, confirmation of the presence of the analytes in samples is possible using the developed methodology¹³⁶.

Table 4.6. MS and MS/MS spectrometry conditions selected for the analysis method.

Mass spectrometry conditions	
Interface	ESI Agilent Jet Stream®
Nozzle	0 V
Fragmentor voltage	75 V
Dry gas	300 °C, 10 L/min
Nebulizer pressure	40 psig
Sheath gas	325 °C, 11 L/min
Vcap	3000 V
Skimmer	65 V
OCT 1 RF	750 V
Mass range	MS: 50-1000 <i>m/z</i> MS/MS: 30-1000 <i>m/z</i>

4.7.3. OPTIMIZATION OF SAMPLE TREATMENT CONDITIONS

The optimization of specific sample treatment procedure in targeted metabolomics methodologies is of great interest for the extraction of analytes as well as for decreasing any matrix effect in order to obtain low limits of quantification.

4.7.3.1. Aminothiols reduction

Sample preparation step should consider the particularities and needs of the analytes of interest. Some amino acids contain thiol groups (aminothiols) and are critical intracellular and extracellular redox buffers that may oxidate forming disulfides (-S-S-), thus binding to proteins or to other aminothiols¹³⁷. Therefore, when measuring aminothiols, it should be taken into account that oxidation and reduction reactions may occur after blood extraction, unless thiol compounds are blocked by derivatization¹³⁸. Besides, it is possible to measure total aminothiol concentration using different reducing agents instead¹³⁹.

Regarding that CYS, GSH and HCYS are aminothiols, there is a need for a reduction step during sample treatment to be able to measure total aminothiols. Otherwise, interconversions between oxidized and reduced species may occur in addition to protein binding by means of disulfide bonds. Moreover, literature showed that the use of reducing agents provides better stability to the sample¹⁴⁰.

For all these reasons, the use of two different reducing reagents was assayed in this study: 2-mercaptoethanol (2-ME) and dithiothreitol (DTT). On the one hand, 5 µL of 2-ME 15 %

was added to 50 μL of plasma and the mixture was incubated for 30 min at 37 $^{\circ}\text{C}$ ¹⁴¹. Then, protein precipitation was performed with 150 μL of acetonitrile and subsequent evaporation and reconstitution in 200 μL of ammonium formate 5 mM was carried out. On the other hand, the addition of 50 μL of DTT 77 g/L to 50 μL of plasma and incubation of 15 min at room temperature followed by the same sample treatment steps was performed¹⁴².

Preliminary studies did not show relevant differences between both aminothiols reduction methods. Therefore, DTT was chosen as it is a simpler procedure, it avoids heating the sample and in comparison with 2-mercaptoethanol, it barely has any odor and the toxicity is lower^{143,144}.

The use of DTT had different effects on the aminothiols. For GSH and CYS, the signal of their protonated ions in untreated plasma was minimal compared to the signal obtained in plasma aliquots treated with DTT. In fact, when using no reducing agents CYS's oxidized form, cystine dimer, was the predominant specie in his protonated form (241.0311 m/z), being CYS signal minimal (122.0270 m/z). In contrast, when treating plasma samples with DTT cystine dimer disappeared completely, thus showing that the reduction process had been completed (see Fig. 4.28).

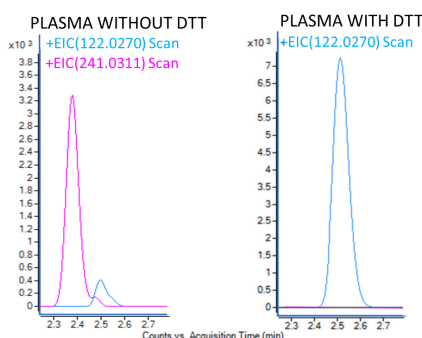


Fig. 4.28. Chromatograms corresponding to two different aliquotes from the same plasma. Plasma sample on the left was not treated with DTT, whereas right chromatogram shows plasma treated with DTT. Cystine homodimer signal is represented in pink colour and CYS signal is coloured blue.

Concerning GSH, no signal of the protonated specie (613.1592 m/z) of the homodimer glutathione disulfide molecule (GSSG) was observed. This might be in relation with poor ionization of this molecule in the selected conditions, taking into account that this molecule is bigger than the analytes of interest and different MS conditions might be necessary to favour its ionization.

Regarding HCYS, no significant differences were found between adding and not adding DTT. In this case, it is suggested that this is related to the fact that around 80-90 % of HCYS in plasma is coupled to proteins⁸⁸, and protein precipitation by itself breaks the bonds between homocysteine and proteins¹⁴⁰.

Furthermore, taking into account that the majority of the metabolites of interest are not aminothiols, we wanted to know the effect that DTT might have on the rest of the analytes. Signals obtained from the same plasma treated with and without DTT were compared and HARG resulted the only analyte for which the signal decreased when using DTT. Nevertheless, the resulting area after sample pretreatment with DTT was still enough to be able to quantify HARG in plasma, as signals in the range of $\mu\text{g/mL}$ were expected (see Fig. 4.29).

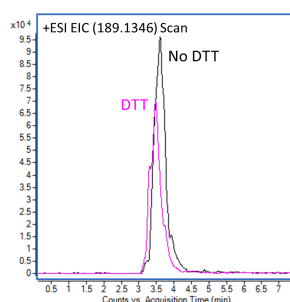


Fig. 4.29. EIC chromatograms of HARG corresponding to two aliquotes from the same plasma in which DTT was added (pink) and not added during sample treatment (black).

4.7.3.2. Sample cleaning and metabolite extraction step

Efficient sample cleaning and metabolite extraction steps are indispensable for reducing matrix effects and guaranteeing the extraction of the metabolites of interest. Two different sample pretreatment approaches were evaluated to achieve a good clean-up of the sample and extraction of the analytes: protein precipitation (PPT) and solid-phase extraction (SPE).

Different solvents were assayed for PPT as precipitation reagents in plasma/precipitant reagent ratios 1:3 to 1:4 based on previous literature works: 1 % formic acid in acetonitrile¹⁴⁵, 0.5 % formic acid in methanol¹⁰⁸, 0.1 N HCl in methanol¹¹⁶ and acetonitrile¹⁴⁶.

The results obtained for each analyte were evaluated as a function of the precipitation reagent used. When using methanol as a precipitation agent, after PPT and extract

evaporation steps a gelatinous residue remained in the bottom of the chromatographic vial, whereas acetonitrile provided cleaner samples. For that reason, the use of methanol was discarded. Regarding the use of acetonitrile alone or containing 1% formic acid, more reproducible results were obtained with acetonitrile alone. Therefore, the use of acetonitrile was selected as precipitation reagent and this solvent was kept at -40°C until PPT in order to favour stability of the analytes.

In addition, after PPT, different SPE cartridges were assayed with the aim of getting cleaner plasma extracts and removing interferences: Oasis MCX 3 cc (60 mg) and Hybrid SPE Phospholipid Ultra 30 mg/1 mL cartridges.

Oasis MCX cartridges contain a mix mode polymeric sorbent, optimized to extract basic compounds with functional groups enabling strong cation exchange with high sensitivity and selectivity. This cartridge was assayed using the following procedure, based on Armstrong et al¹⁰⁷:

1. Sorbent activation was carried out adding 1 mL of methanol.
2. SPE was conditioned using 1 mL of HCl 0.1 N.
3. Plasma extract was loaded.
4. The addition of 1 mL of methanol was performed, as a cleaning step.
5. Elution was carried out by means of 1 mL of methanol solution containing 5 % of ammonia.

Finally, the eluate was evaporated in N_2 stream and reconstituted in 5 mM ammonium formate.

Plasma extracts obtained from PPT alone and from PPT followed by the procedure described above for SPE Oasis MCX cartridge were analyzed and compared. According to these results, this cartridge does not perform appropriately for GSH, HCYS and SAH, for which the recuperation is not enough to allow their quantification in plasma.

Secondly, Hybrid SPE Phospholipid Ultra cartridge was tested, which contains zirconium-coated silica that interacts with phospholipids as an electron acceptor. This results in phospholipid binding to the surface of the cartridge, reducing matrix effect associated to these compounds. After the corresponding protein precipitation step, plasma extract was passed through the cartridge with the aim of verifying whether phospholipid elimination could decrease the observed matrix effect.

However, after comparing the use of PPT alone with the addition of Hybrid SPE clean-up step, results showed that this SPE procedure was not appropriate for GSH, CYS, MET, ADMA, SDMA, HCYS, SAH and SAM, which kept retained in the cartridge.

Therefore, regarding that the recovery was not good enough using neither of the cartridges, SPE was discarded, and PPT was selected as the unique technique for extraction.

Since some of the analytes were in concentrations above the linearity range, the election of a dilution factor of the sample is indispensable. Indeed, it has to be noted that for some analytes, such as, CNN, CTN, ARG and CYS, concentrations in the range of 10-30 µg/mL were expected in plasma according to preliminary studies. However, some other analytes like SAH and SAM were expected in the range of a few ng/mL, which means that the dilution factor should allow the detection and quantification in plasma of these metabolites.

Three different dilution factors were assayed, 1:2, 1:4 and 1:10. For that purpose, the following procedure was adopted: 50 µL of plasma sample followed PPT, evaporation in nitrogen stream and the reconstitution of the sample in the correspondent 5 mM ammonium formate volume (100 µL, 200 µL or 500 µL).

Finally, reaching a compromise between obtaining the major reduction of matrix effect and at the same time allowing the detection of minority analytes, 1:4 dilution factor was selected. Thus, 50 µL of plasma were used for sample treatment and after protein precipitation, samples were evaporated and subsequently reconstituted in 200 µL of ammonium formate 5 mM.

4.7.3.3. Addition of isotopically labeled compounds

Quantification of analytes must be performed with the guarantee that the steps followed during sample treatment do not affect the final result as a consequence of differences in metabolite extraction. For that reason, isotopically labeled compounds were added as internal standards to correct for those minor differences suffered by the samples. Despite the fact that correcting the signal of each analyte with its corresponding isotopically labeled analogue would be the ideal, it is economically unfeasible.

For that reason, in this case, four different isotopically labeled compounds were used: glutathione-¹³C₂¹⁵N (GSH-¹³C₂¹⁵N), creatine-d₃ (CTN-d₃), creatinine-d₃ (CNN-d₃) and

symmetric dimethylarginine- d_6 (SDMA- d_6). These isotopically labeled compounds were selected because they covered different retention times in the chromatogram, thus enabling better signal correction. In addition, GSH- $^{13}C_2^{15}N$ is an aminothiols and would be useful to check the performance of these sulfur containing amino acids, and SDMA- d_6 allows its MS and MS/MS analysis for monitorization of the structural isomers quantified in MS/MS mode.

Accordingly, CTN- d_3 corrected the signal of GLY, CIT, DMG, BET, CYS and CTN, whereas GSH- $^{13}C_2^{15}N$ was used to adjust GSH signal. Similarly, CNN- d_3 was used to correct for the concentration of HCYS, ARG, MET, CNN and HARG, and SDMA- d_6 for ADMA, SDMA, SAM and SAH adjustment. SDMA- d_6 was analyzed both in MS and in MS/MS modes and signal from MS mode was used to correct SAM and SAH signal, whereas MS/MS adjusted for ADMA and SDMA. 10 μ L of a mix solution containing 75 μ g/mL of CNN- d_3 , CTN- d_3 and GSH- $^{13}C_2^{15}N$ and 25 ng/mL of SDMA- d_6 were added to 50 μ L of plasma, to reach concentrations of 15 μ g/mL and 5 ng/mL respectively.

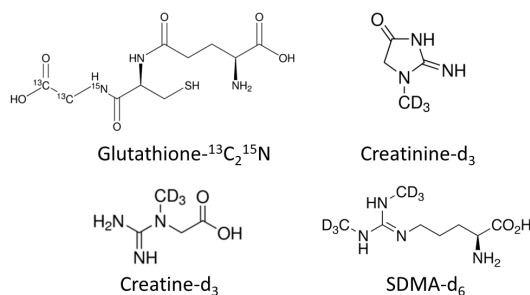
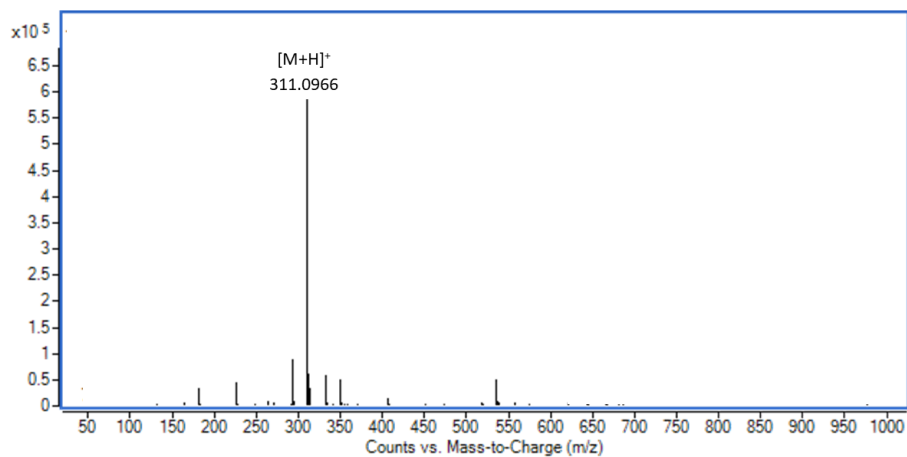
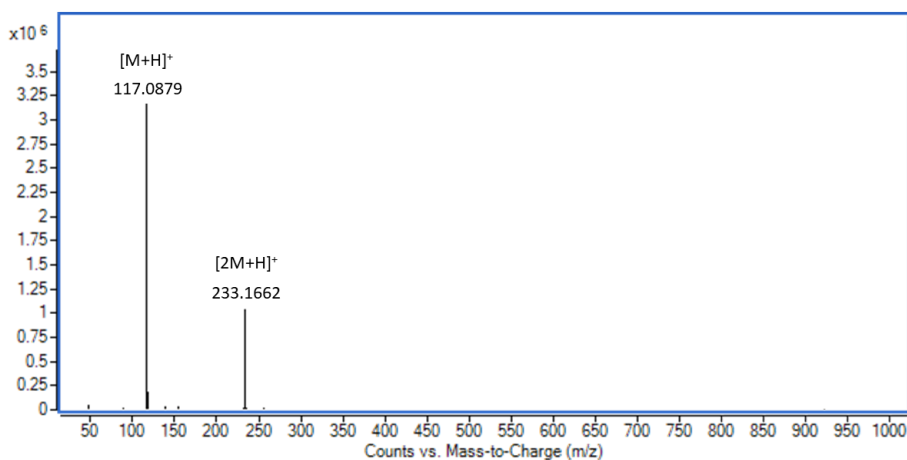
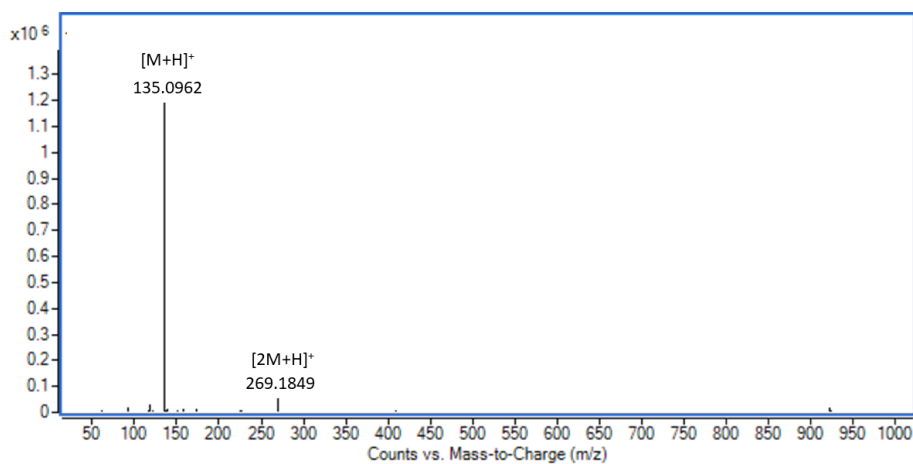
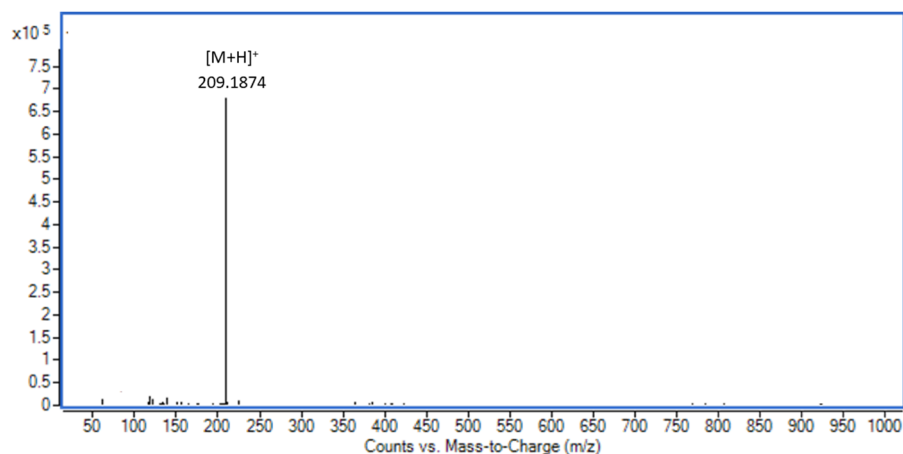
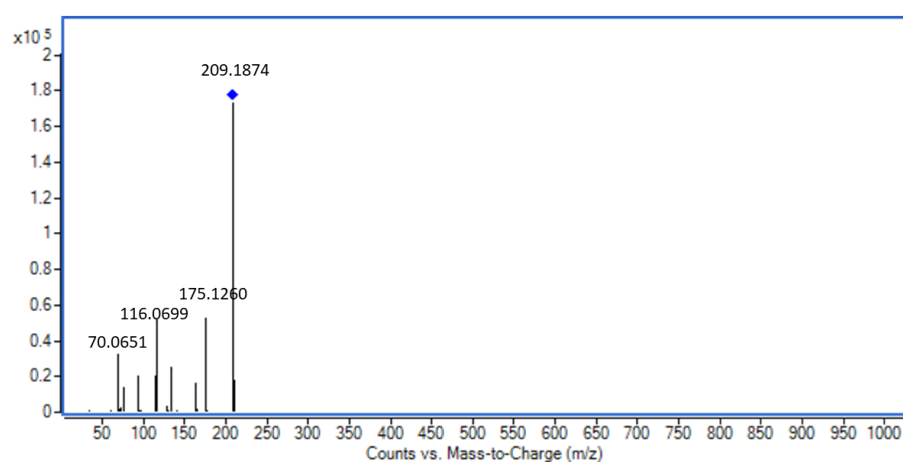


Fig. 4.30. Chemical structure of isotopically labeled compounds used for signal correction.

Protonated species of these isotopically labeled compounds were used for signal correction in MS mode: 135.0956 m/z for CTN- d_3 , 311.0955 m/z for GSH- $^{13}C_2^{15}N$, 117.0850 m/z for CNN- d_3 and 209.1879 m/z for SDMA- d_6 . For SDMA- d_6 , MS/MS transition 209.1879 > 175.1274 m/z , corresponding to $[M+H]^+ > [M-CHD_3N]$, was used for SDMA and ADMA corrections. MS spectra of 5 μ g/mL of these four isotopically labeled compounds as well as MS/MS spectra of SDMA- d_6 are showed in Fig. 4.31.

✓ **GSH- $^{13}\text{C}_2^{15}\text{N}$ (MS)**✓ **CNN-d₃ (MS)**✓ **CTN-d₃ (MS)****Fig. 4.31.** MS and MS/MS spectra of the isotopically labelled compounds.

✓ **SDMA-d₆ (MS)**✓ **SDMA-d₆ (MS/MS)****Fig. 4.31 (cont.).** MS and MS/MS spectra of the isotopically labelled compounds.

Taking into account all of these concerns, sample treatment consisted on the following procedure. First, plasma samples were thawed and 50 μL of plasma were placed in an Eppendorf tube. Then, internal standard addition was performed by spiking 10 μL of a mix solution to reach concentrations of 15 $\mu\text{g}/\text{mL}$ creatinine-d₃, creatine-d₃ and glutathione-¹³C₂¹⁵N, and 5 $\mu\text{g}/\text{mL}$ of SDMA-d₆. Afterwards, 50 μL of dithiothreitol (77 g/L) were added and incubated for 15 min at room temperature for the reduction of aminothiol compounds. 150 μL of cold acetonitrile were added and vortexed to perform plasma protein precipitation. Then, samples were centrifuged at 13000 rpm for 10 min at 4°C and the supernatants were transferred to chromatography vials. These supernatants were evaporated in nitrogen stream and finally reconstituted in 200 μL of ammonium formate 5 mM. It has to be noted that each sample was extracted in triplicate.

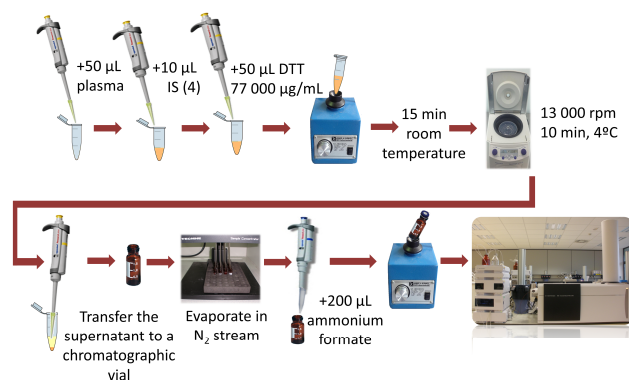


Fig. 4.32. Summary of the sample treatment procedure carried out for plasma samples.

4.8. ANALYTICAL METHOD EVALUATION

4.8.1. CALIBRATION OF THE METHOD AND MATRIX EFFECT

Quantification of analytes in a matrix is ideally performed by means of an analyte-free matrix spiked with known quantities of the analytes of interest. Then, a calibration curve is built and signals obtained for real samples are interpolated using the equation of the curve. However, in quantification of endogenous compounds it is really difficult to obtain analyte-free matrices.

In this case, there is not any plasma commercialized that does not contain any of the analytes of interest, thus making impossible the quantification by means of the addition of analytical standards to a blank matrix at known quantities. Different alternatives may be considered: matrices with depreciable levels of analytes (e.g. quantification of female hormones using male plasma), *in vivo* or *in vitro* methods to reduce their concentration in biological matrix or surrogate matrices¹⁴⁷. Once considered the advantages and inconveniences of these options and regarding that the analytes of interest had a varied nature, the use of a surrogate matrix was considered the best possibility to obtain a matrix free of analytes.

First of all, the feasibility of preparing analytical standards in 5 mM ammonium formate was considered. Comparison of the matrix effect of ammonium formate and plasma was performed using the four isotopically labeled compounds described in section 4.7.3.3 at two concentration levels and areas obtained were compared. Low concentration level contained 1.5 µg/mL CNN-d₃, CTN-d₃ and GSH-¹³C₂¹⁵N and 50 ng/mL of SDMA-d₆, whereas

high concentration level contained 3 $\mu\text{g/mL}$ CNN-d₃, CTN-d₃ and GSH-¹³C₂¹⁵N and 100 ng/mL of SDMA-d₆. Matrix factor (MF) was calculated dividing the result of the spiked analyte in plasma with the result of the analyte spiked in ammonium formate.

Table 4.7. Plasma vs. ammonium formate 5 mM matrix factor calculation.

MF (%)	Plasma / Ammonium formate 5 mM	
	Low level	High level
CNN-d ₃	58.8 ± 1.1	66.3 ± 1.2
CTN-d ₃	18.5 ± 0.3	26.0 ± 0.2
GSH- ¹³ C ₂ ¹⁵ N	95.5 ± 3.2	85.5 ± 1.2
SDMA-d ₆	83.0 ± 8.2	91.1 ± 0.2

Even though for GSH-¹³C₂¹⁵N and SDMA-d₆ similar areas were obtained in plasma and 5 mM ammonium formate, different matrix effects were found CTN-d₃ and CNN-d₃. Indeed, CTN-d₃ area was more than 4 times higher in ammonium formate 5 mM in comparison with plasma at the same concentration. Similarly, CNN-d₃ area in ammonium formate doubled its area in plasma.

Regarding that calibration curve could not be prepared in ammonium formate 5 mM because of the matrix effects, alternatively, Serasub® serum surrogate matrix was assayed. Serasub® is characterized for being a protein free liquid in a buffered solution (pH 7.4), equivalent to serum and plasma in terms of specific gravimetry, viscosity and osmolality. As in previous case, the same two concentration levels were used to obtain the results of plasma/Serasub® ratio expressed as a percentage (Table 4.8).

Table 4.8. Plasma vs. Serasub® matrix factor calculation.

MF (%)	Plasma / Serasub®	
	Low level	High level
CNN-d ₃	71.8 ± 0.1	81.4 ± 1.7
CTN-d ₃	32.9 ± 0.9	44.8 ± 2.2
GSH- ¹³ C ₂ ¹⁵ N	141.0 ± 24.7	131.1 ± 19.7
SDMA-d ₆	90.9 ± 2.6	100.2 ± 3.0

Again, significant differences were found for areas of different analytes at the same concentration in plasma and Serasub®. Indeed, except for SDMA-d₆, for which areas are quite similar for these three matrices, the rest of the analytes had differing areas and showed to be affected by the nature of the matrices. Moreover, GSH-¹³C₂¹⁵N presented a higher area in plasma, whereas CNN-d₃ and CTN-d₃ showed lower area in plasma in

comparison with Serasub®. From these results it can be inferred that their correspondent non-isotopically modified analogs will present different areas in Serasub® synthetic serum in comparison with plasma matrix as well.

Consequently, ammonium formate and Serasub® were discarded for quantification and standard addition quantification in plasma matrix was considered the best alternative. However, sample volume was not enough to carry out as many additions as necessary per plasma sample for a good quantification. Therefore, a plasma pool was made prepared from patients suffering from CKD and control samples. This pool was quantified by standard addition and was used as the first level of the calibration curve. In addition, it was spiked with known concentrations of the analytes of interest to obtain the rest of the calibration levels. All the calibration levels were subjected to the same sample treatment process developed for individual plasma samples and analyzed in duplicate. Signals of the analytes were corrected with their corresponding internal standards for each calibration level, as showed in section 4.7.3.3. Linear regression equations of the analytes of interest were obtained facing the relation between analytes and internal standard signal *versus* analyte concentration (see Table 4.9). In this case, correlation coefficients of at least 0.9936 were achieved.

Table 4.9. Calibration curves and statistical analysis on the random distribution of the residues for each analyte of interest in pooled plasma.

Analytes	Calibration range (µg/mL)	Slope ± error	Intercept ± error	Correlation coefficient (r)	σ_{TOT}	\bar{r}	$\sigma(r)$	$ \bar{r} $
ADMA	0.159-1.4	11.77 ± 0.75	0.78 ± 0.20	0.9939	0.16	0.00	0.13	0.11
ARG	14.3-46.8	1.05 ± 0.06	7.75 ± 1.88	0.9967	1.45	0.00	1.02	0.89
BET	3.1-20.6	3.42 ± 0.10	9.05 ± 1.10	0.9991	1.45	0.00	1.02	0.81
CIT	2.8-17.8	0.50 ± 0.00	0.57 ± 0.07	0.9999	0.11	0.00	0.08	0.07
CTN	9.7-42.2	0.29 ± 0.01	1.58 ± 0.12	0.9999	0.11	0.00	0.08	0.06
CNN	15.2-30.2	0.30 ± 0.02	-0.17 ± 0.30	0.9949	0.20	0.00	0.16	0.13
CYS	13.6-111.1	0.20 ± 0.01	1.29 ± 0.76	0.9969	0.82	0.00	0.58	0.54
DMG	1-17.3	1.04 ± 0.03	0.70 ± 0.25	0.9984	0.42	0.00	0.35	0.29
GSH	3.2-19.5	0.21 ± 0.01	0.57 ± 0.05	0.9988	0.05	0.00	0.03	0.03
GLY	1.8-10.5	1.67 ± 0.05	0.15 ± 0.57	0.9990	0.54	0.00	0.38	0.31
HARG	0.4-3.7	5.78 ± 0.41	1.42 ± 0.54	0.9951	0.54	0.00	0.38	0.31
HCYS	0.5-8.7	4.85 ± 0.23	2.74 ± 0.41	0.9967	0.62	0.00	0.48	0.38
MET	2.7-11.5	0.31 ± 0.01	0.40 ± 0.04	0.9995	0.04	0.00	0.03	0.02
SAH	0.036-0.524	48.82 ± 0.92	0.57 ± 0.08	0.9996	0.08	0.00	0.06	0.04
SAM	0.055-0.555	78.89 ± 2.26	2.13 ± 0.44	0.9992	0.05	0.00	0.03	0.03
SDMA	0.291-1.500	2.74 ± 0.22	0.29 ± 0.09	0.9936	0.08	0.00	0.06	0.04

It is essential to verify that the linear regressions are adequately adjusted to data obtained by means of a statistical analysis on the random distribution of the residues. To comply with this assumption, it was verified that standard deviation of the residues, $\sigma(r)$, was equal to or less than global standard deviation, σ_{TOT} , and average residue, $|\bar{r}|$, was equal to or lower than $\sigma(r)$. In addition, if data distribution is normal, the arithmetic mean of the residues, \bar{r} , is expected to be close to 0.

The results of the statistical analysis on the random distribution of the residues are showed in Table 4.9. As it can be observed, all the assumptions explained are fulfilled. Thus, it is concluded that the equations of linear regression are adequate to quantify plasma samples.

Taking into account that concentrations below the first calibration level were expected for some real samples, it was considered necessary to verify the extent of the linearity of the calibration curve obtained, to consider whether it might be possible the extrapolation of lower concentrations with some degree of confidence. Isotopically labeled compounds were used to check the linearity of the method in a defined calibration range and lower concentration levels were introduced until reaching the lower limit of quantification (LLOQ), the lowest concentration of analyte in a sample that can be determined or quantitated with acceptable relative standard deviation.

The method was linear between 100 ng/mL and 50 $\mu\text{g/mL}$ for GSH- $^{13}\text{C}_2^{15}\text{N}$, 5 ng/mL to 20 $\mu\text{g/mL}$ for CNN- d_3 , 5 ng/mL to 50 $\mu\text{g/mL}$ for CTN- d_3 and 2 ng/mL to 50 $\mu\text{g/mL}$ for SDMA- d_6 in MS mode. Regarding calibration curves and correlation coefficients obtained for isotopically labeled compounds, these are showed in Fig. 4.33. As it can be observed, correlation coefficient is above 0.9969 in all cases.

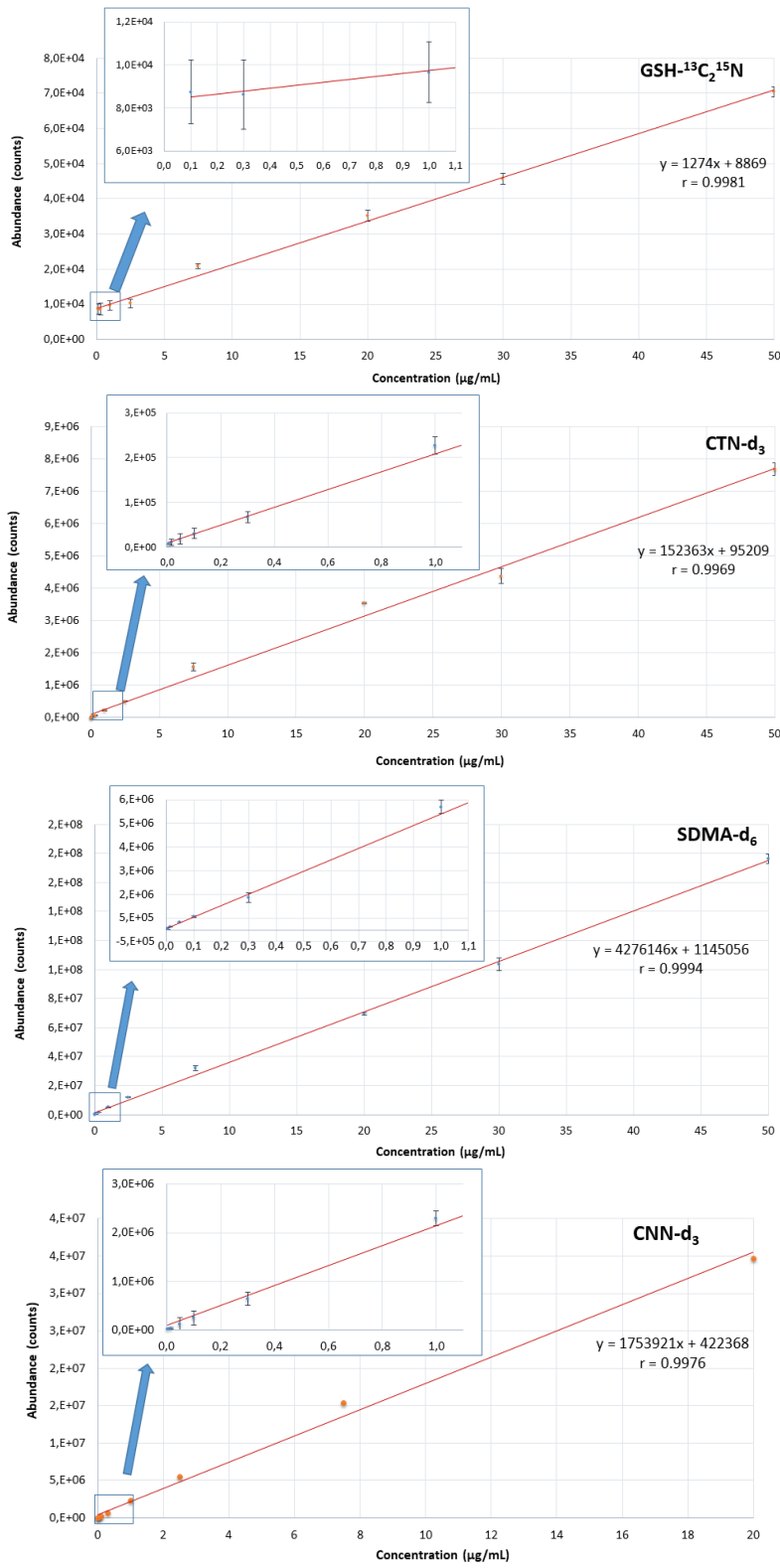


Fig. 4.33. Calibration curves of isotopically labeled compounds in MS mode.

4.8.2. LIMITS OF QUANTIFICATION, ACCURACY AND PRECISION

The sensitivity of this LC-QTOF-MS method in plasma is defined by the lower limit of quantification (LLOQ). Regarding that there was not any commercialized plasma free of the analytes of interest, decreasing concentrations of isotopically labeled compounds were added and included in a calibration curve, being the first concentration level of the calibration curves set as LLOQ. Based on this, LLOQ were determined as follows: 100 ng/mL for GSH in MS mode, 5 ng/mL for CTN and CNN in MS mode, 2 ng/mL for SDMA in MS mode and 5 ng/mL for SDMA in MS/MS mode. These LLOQ were compared with those from recent analytical methods from literature in order to evaluate the contribution of the method. The LLOQ for CTN and GSH were found to be similar to those from recent works dealing with LC-MS in plasma matrix^{148,149}, while the LLOQ for CNN and SDMA were 5 and 2-fold lower, respectively^{150,151}.

Both accuracy and precision of the developed method were calculated for unspiked, low, medium and high QC samples (described in section 4.6.3), performing five determinations per level and repeating them for 3 days.

As it can be observed in Table 4.10, both accuracy and precision fell within the acceptable range in this study. Indeed, accuracy was between 83.2 % and 116.3 % in all cases, whereas, the relative standard deviation (RSD) of repeated individual measures was below 13.7 % for all the analytes. Likewise, intermediate precision was below 16.7 % for all the compounds.

Table 4.10. Summary of the accuracy and precision results of the LC-QTOF-MS method for the analyzed quality control samples.

	Within-run accuracy (%nominal value)				Within-run precision (%RSD)				Intermediate precision (%RSD)			
	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC
ADMA	96.3	100.3	100.6	99.0	5.8	1.4	1.0	4.9	12.4	9.5	4.7	10.0
ARG	111.8	99.8	89.6	85.2	1.3	3.6	1.4	4.8	6.1	3.5	5.6	5.0
BET	106.3	94.8	90.5	116.3	2.0	3.5	1.4	3.1	7.2	6.6	7.6	5.1
CIT	98.8	97.0	83.7	95.7	1.9	5.7	3.6	3.4	6.8	4.0	5.1	2.6
CTN	93.8	93.7	99.0	100.5	1.7	1.2	1.3	1.7	7.6	12.9	10.1	9.2
CNN	100.2	96.7	89.3	92.6	0.8	0.4	0.5	1.2	0.6	0.3	0.3	1.2
CYS	110.9	97.6	83.6	97.7	4.2	1.2	9.2	3.6	13.7	14.4	14.4	16.0
DMG	113.7	101.5	88.1	98.3	2.0	7.1	3.9	3.5	16.2	8.8	9.0	6.8
GSH	103.2	106.4	102.0	108.4	6.1	10.0	7.1	4.7	1.4	3.8	7.9	1.5
GLY	111.6	100.9	83.2	93.8	3.6	13.7	4.1	5.5	4.5	11.0	10.5	9.4
HARG	95.1	96.7	92.0	96.4	1.0	5.5	1.9	6.0	10.4	14.5	12.3	14.7
HCYS	91.2	98.6	100.7	102.5	5.7	3.4	2.6	4.2	5.3	0.9	4.8	4.5
MET	112.7	97.5	101.5	100.1	1.2	3.7	3.7	5.7	9.4	9.1	3.3	4.1
SAH	102.3	94.0	97.3	97.0	8.6	4.7	1.3	2.7	14.4	10.1	14.9	13.0
SAM	95.4	101.3	99.7	109.7	4.9	1.6	2.0	2.2	8.4	13.5	16.7	13.7
SDMA	101.8	110.2	101.8	98.8	7.7	11.0	10.1	9.3	9.5	6.4	8.6	10.1

4.8.3. RECOVERY FROM THE EXTRACTION PROCESS

The efficiency of sample preparation procedure is usually evaluated by means of recovery assessment. Regarding that, analyte-free matrix was not available, addition of isotopically labeled compounds before and after the extraction process at two concentration levels in unspiked QC samples was used to evaluate the recovery of the extraction process. First level consisted in the addition of isotopically labeled compounds to obtain concentrations of 1.5 µg/mL of CNN-d₃, CTN-d₃ and SDMA-d₆, and 100 ng/mL of GSH-¹³C₂¹⁵N, whereas second level corresponded exactly to the double of first level.

Table 4.11. Recovery and standard deviation of isotopically labeled analytes in two concentration levels.

Recovery (%)	Low level	High level
CNN-d ₃	86.8 ± 1.5	107.4 ± 0.4
CTN-d ₃	83.9 ± 3.0	92.5 ± 2.3
GSH- ¹³ C ₂ ¹⁵ N	74.9 ± 10.8	82.1 ± 8.0
SDMA-d ₆	85.5 ± 12.3	97.5 ± 7.8

Taking into account the recoveries observed in Table 4.11, which are within 82.1-107.4 % range, it can be concluded that the developed sample treatment procedure provides a good recovery for the analytes studied.

4.8.4. STABILITY

Stability of analytes needs to be assessed in order to obtain high quality data from plasma sample analysis, that is, data not affected by any degradation process that might occur in analytical standards or in plasma. For that purpose, areas of the analytes analyzed before and after subjecting analyte containing matrices to different conditions, such as, room temperature, freezing, and freezing and thawing cycles, were compared. Table 4.12 shows detailed data of stability studies carried out.

Table 4.12. Stability of the analytes in stock solutions, plasma samples and plasma extracts.

COMPOUND	Stock solutions						Plasma samples			Plasma extracts		
	Ambient temperature 48 h	Frozen 1 week	Frozen 1 month	1 freeze and thaw cycle	2 freeze and thaw cycles	1 freeze and thaw cycle	1 freeze and thaw cycle	Ambient temperature 24 h	Frozen 1 week	1 freeze and thaw cycle	Frozen 1 week	1 freeze and thaw cycle
ADMA	99.0 ± 0.7	97.2 ± 4.3	86.7 ± 2.3	98.4 ± 1.4	100.6 ± 0.9	98.0 ± 36.1	114.7 ± 9.16	112.1 ± 0.7	100.9 ± 1.19			
ARG	89.0 ± 8.3	93.8 ± 5.5	113.6 ± 4.7	94.1 ± 5.3	97.3 ± 7.7	108.0 ± 7.2	99.3 ± 18.7	87.6 ± 31.2	97.9 ± 6.2			
BET	99.5 ± 0.3	99.1 ± 1.3	102.0 ± 1.9	101.1 ± 0.8	97.0 ± 1.8	90.0 ± 32.6	104.0 ± 2.2	91.4 ± 44.6	88.5 ± 9.9			
CIT	101.2 ± 0.8	85.9 ± 0.8	114.2 ± 0.8	94.4 ± 1.0	107.2 ± 1.2	98.5 ± 10.4	109.9 ± 1.9	94.8 ± 0.8	99.4 ± 1.0			
CTN	99.0 ± 0.8	98.1 ± 2.7	105.0 ± 3.5	97.5 ± 4.2	93.2 ± 1.6	94.6 ± 45.3	105.7 ± 5.0	96.3 ± 13.0	97.9 ± 5.0			
CNN	90.3 ± 7.2	97.0 ± 0.9	104.1 ± 1.2	101.9 ± 4.1	104.8 ± 2.0	90.0 ± 12.3	108.9 ± 3.9	87.9 ± 3.5	98.3 ± 3.3			
CYS	96.2 ± 2.8	98.3 ± 1.3	97.3 ± 10.2	124.8 ± 14.9	85.0 ± 5.0	12.8 ± 0.3	57.8 ± 4.6	88.9 ± 10.0	98.5 ± 2.8			
DMG	94.4 ± 4.1	101.7 ± 8.8	61.7 ± 2.3	85.0 ± 7.1	107.5 ± 1.6	89.7 ± 4.9	113.9 ± 17.1	95.9 ± 34.7	98.8 ± 10.8			
GSH	101.8 ± 1.3	103.8 ± 22.0	119.8 ± 14.5	104.9 ± 4.9	71.2 ± 15.1	93.4 ± 1.7	89.4 ± 14.0	102.4 ± 6.0	87.7 ± 10.7			
GLY	88.6 ± 11.0	98.7 ± 6.7	53.9 ± 7.8	113.0 ± 3.0	85.3 ± 5.7	88.8 ± 0.4	96.3 ± 1.3	86.0 ± 11.7	104.8 ± 4.6			
HARG	94.0 ± 4.4	111.2 ± 2.1	86.1 ± 1.4	86.7 ± 3.0	106.8 ± 15.7	94.7 ± 9.6	99.3 ± 1.3	87.1 ± 37.3	88.7 ± 10.7			
HCYS	94.0 ± 4.4	114.3 ± 3.4	114.4 ± 8.7	124.3 ± 7.4	123.0 ± 4.7	98.5 ± 6.7	101.5 ± 11.1	97.1 ± 12.9	98.2 ± 3.9			
MET	89.8 ± 7.6	103.2 ± 3.6	121.4 ± 5.9	95.4 ± 6.1	95.0 ± 12.4	103.5 ± 1.9	98.0 ± 8.3	99.7 ± 1.0	98.6 ± 2.6			
SAH	101.3 ± 0.9	95.6 ± 2.1	95.6 ± 4.5	87.4 ± 1.0	70.7 ± 2.3	98.9 ± 8.8	102.1 ± 2.9	90.6 ± 21.4	97.9 ± 5.1			
SAM	90.8 ± 6.8	82.4 ± 3.6	54.3 ± 1.2	118.5 ± 13.0	111.4 ± 2.0	100.3 ± 0.1	115.4 ± 16.0	122.1 ± 35.7	97.9 ± 7.5			
SDMA	98.9 ± 0.8	105.2 ± 3.0	86.8 ± 2.5	104.8 ± 2.4	95.9 ± 8.7	118.2 ± 19.2	115.4 ± 3.5	108.0 ± 14.7	97.9 ± 4.8			

Regarding individual stock solutions, they proved to be stable frozen for more than a month. However, combined stock solutions were less stable. Indeed, these solutions were stable for at least 48 h at room temperature (89–101 %) and frozen for at least 1 week (82–114 %), but less than a month (54–121 %). For that reason, combined stock solutions were weekly prepared from individual stock solutions.

In addition, plasma extracts showed to be stable at room temperature for 24 h (89–115 %), except for CYS, which was stable for 12 h. Nevertheless, plasma extracts proved to be more stable frozen, as they were stable for more than 1 week frozen (86–122 %), even though multiple freeze and thaw cycles affects their stability. Besides, the effect of freezing and thawing plasma samples before sample treatment was also assayed, being observed stability (89–105 %) for all the metabolites except for CYS.

4.9. SAMPLE ANALYSIS

Once the analytical method was validated, the analysis of real plasma samples was performed. Calibration standards made by spiking plasma, as previously explained in section 4.8.1, were analyzed twice, in addition to CKD and control extracts which were randomly analyzed. Besides, during the batch QC samples were used to verify the correct performance of the analysis.

An example chromatogram of a plasma sample from a CKD patient is showed in Fig. 4.34.

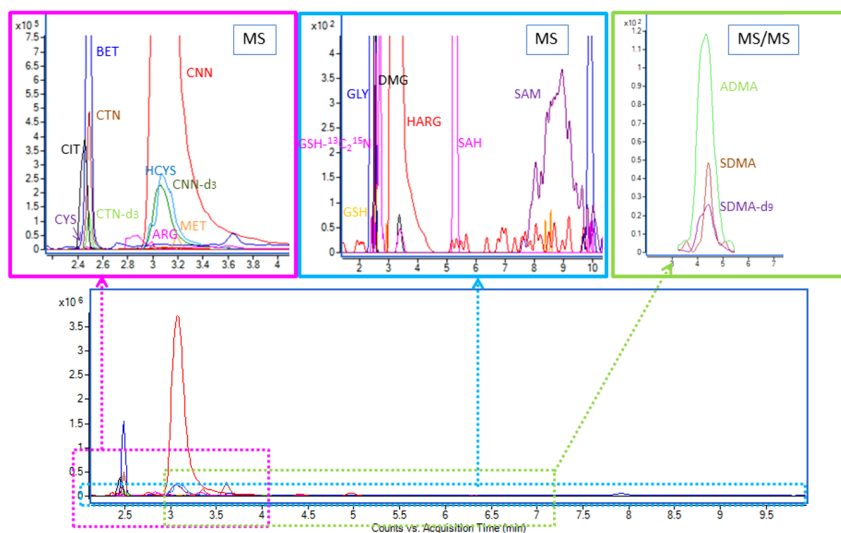


Fig. 4.34. Chromatogram of a plasma sample containing extracted ion chromatograms in MS/MS mode for ADMA and SDMA, and in MS mode for the rest of the analytes.

Quantification of each analyte in plasma was carried out dividing the area of analytes with the area of internal standards and introducing that ratio into the equation of their corresponding calibration curve. An image showing the results of quantification according to a color scale for each analyte and each sample has been made (Fig. 4.35).

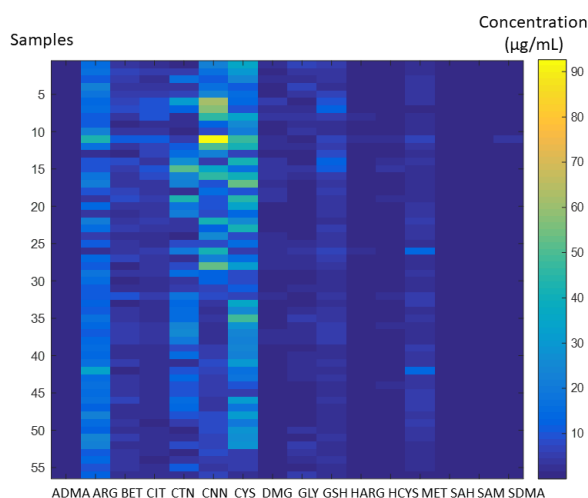


Fig. 4.35. Image graph showing concentration of different analytes *versus* real control and plasma samples in a colour scale.

Concentration ranges vary quite a lot between different analytes as well as between different samples, as it can be observed in this figure. For instance, for CNN, samples in the range of 80-90 µg/mL are found in contrast to other samples below 10 µg/mL. Due to the fact that analytes are also in different concentration ranges, this image graph was rescaled from 0 to 90 µg/mL to 10-50 µg/mL (Fig. 4.36).

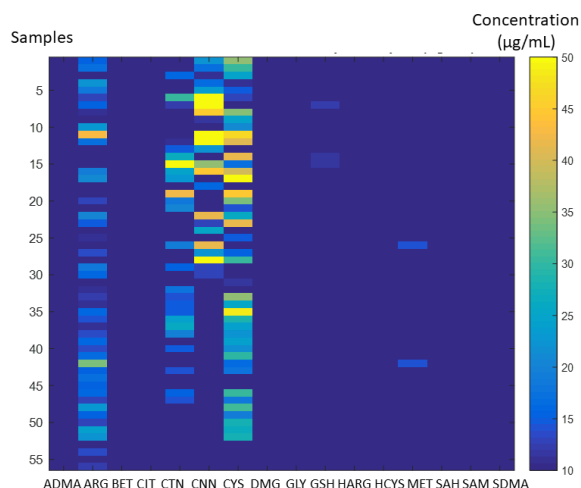


Fig. 4.36. Image graph rescaled to 10-50 µg/mL showing concentration of different analytes *versus* real control and plasma samples in a colour scale.

This rescaled image graph evidences that ARG, CTN, CNN and CYS are in a concentration range from approximately 10-50 $\mu\text{g/mL}$, and that between sample differences are noticed as they show different colours in the colour scale. This image was again rescaled to examine the concentrations of metabolites in a lower concentration range, 0-10 $\mu\text{g/mL}$ (see Fig. 4.37).

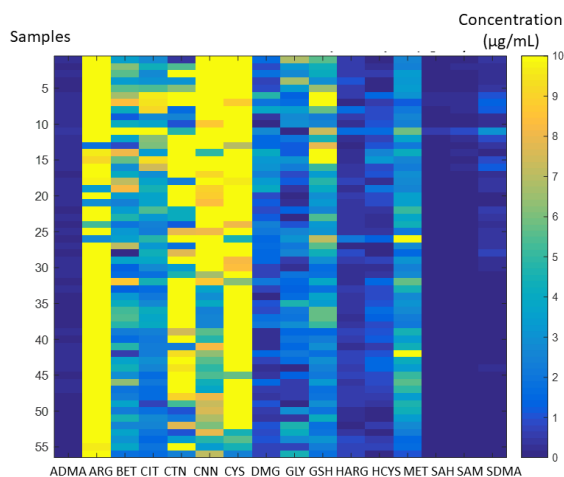


Fig. 4.37. Image graph rescaled to 0-10 $\mu\text{g/mL}$ showing concentration of different analytes *versus* real control and plasma samples in a colour scale.

Having a detailed look at Fig. 4.37, another group with intermediate levels of concentrations is found: BET, CIT, DMG, GLY, GSH, HARG, HCYS and MET. In addition, there are still other metabolites like ADMA, SAH, SAM and SDMA which are found in a lower concentration range, as showed in the image graph rescaled to 0-1 $\mu\text{g/mL}$ (Fig. 4.38).

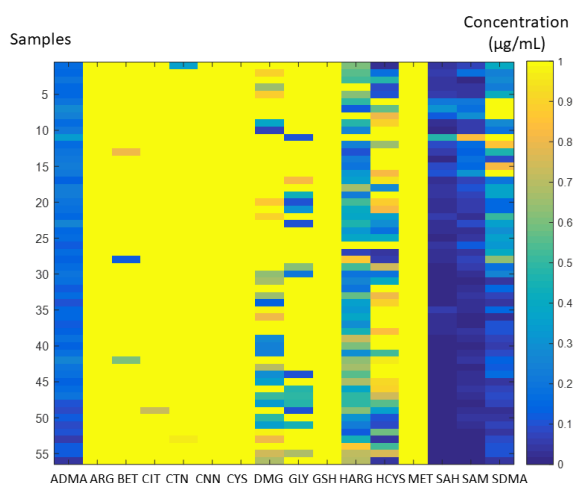


Fig. 4.38. Image graph rescaled to 0-1 $\mu\text{g/mL}$ showing concentration of different analytes against real control and plasma samples in a colour scale.

4.10. DATA ANALYSIS

Once samples are pretreated, analyzed by means of the developed LC-QTOF method and quantified with the calibration curves obtained, there is a need for data analysis. Data analysis makes use of statistics and chemometrics to display significant differences between sample groups and intends to find patterns on data that allow an interpretation of the results from a biological point of view. Data analysis is usually carried out by means of two approaches in targeted metabolomics studies: univariate methods and multivariate methods.

Univariate methods compare different group of samples taking into account a single analyte at a time. Targeted metabolomics studies generally deal with the comparison of median or mean values between different sample groupings. For that purpose, verifying the normality is usually necessary by means of different tests, such as Kolmogorov-Smirnov test. Then, diverse tests might be used in accordance to the characteristics of data, such as: One-way Anova, Student's *t*-test, *U* test in accordance to Mann and Whitney or Kruskal Wallis.

Regarding multivariate methods, they include a variety of approaches, like unsupervised methods (for instance, Principal Component Analysis –PCA- and clustering), supervised techniques (such as Partial Least Square Discrimination Analysis , PLS-DA) and variable selection and classification methods (e.g. Sparse Partial Least Square Discrimination Analysis, SPLS-DA). These methods are useful for data mining, interpretation of complex data matrices and comparison of different data sets.

With that aim, first of all descriptive statistics were used to have a general idea of data. Then, PCA was carried out on logarithm transformed, autoscaled data by means of Matlab R2015a (Mathworks, Natick, Massachusetts, United States) software. In addition, several multivariate classification methods were compared in their predictive ability for CKD, being Maximum Likelihood – Linear Discriminant Analysis (ML-LDA)¹⁵², Fisher-Linear Discriminant Analysis (Fisher-LDA)¹⁵³, Quadratic Discriminant Analysis (QDA)¹⁵⁴ and *k* Nearest Neighbours (kNN)¹⁵⁵. The accuracy of each classification method was achieved by means of a random division of samples to obtain 80 % of samples in a training set and 20 % of the samples in a test set, and optimization of the parameters that influence the model was performed by means of leave-one-out approach (LOO) application in the training set. For kNN, the optimization of the number of neighbours to be used was necessary, which was done by testing different *k* numbers of neighbours from 1 to 10

using LOO on the training set. Random division of the samples was repeated 50 times and the mean performances were obtained for parameters to be optimized to select the most appropriate. Accordingly, to compare the results from different methods, test set classification based on models built in the training sets were repeated 50 times and mean performances were achieved as well.

Besides, PLS-DA¹⁵⁶ was applied after optimizing the number of latent variables (LV), and SPLS-DA¹⁵⁷ after selecting the number of metabolites to be included. In the case of SPLS-DA the number of latent variables was selected according to the number of groups minus one, as it is advised for generating the most stable models and variable selection was carried out for the election of the most meaningful metabolites¹⁵⁸.

Finally, to judge whether the 15 metabolites measured improve the performance and add to the CNN-based diagnosis, the optimized outcomes of the classifiers were compared to the diagnostic performance by CNN alone, the classical biomarker for CKD.

4.10.1. DESCRIPTIVE STATISTICS

Data analysis of the results obtained from CKD and control patients' plasma started with descriptive statistics, with the aim of having a general idea of the data. First of all, results from quantification of the metabolites were summarized in Table 4.13 showing median and interquartile range (IQR).

Distributions of analyte concentrations in control and CKD sample groups were studied by means of Kolmogorov-Smirnov test, a non-parametric test of the equality of continuous one-dimensional probability distributions used to compare the distribution of a sample group with a reference theoretical distribution or distribution between two sample groups. This test is used to assess which variables are under normal distribution for later application of other parametric or non-parametric tests. DMG, BET and ADMA presented a parametric distribution, whereas GSH, GLY, CYS, CIT, CTN, MET, ARG, CNN, HARG, HCYS, SDMA, SAH and SAM showed non-parametric distribution.

Table 4.13. Results obtained for the analytes of interest using the developed method in plasma expressed as median (3rd-97th interquartile range).

COMPOUND	Control (µg/mL)	CKD (µg/mL)
ADMA	0.133 (0.046-0.217)	0.180 (0.117-0.275)
ARG	14.7 (12.5-16.2)	13.3 (9.9-18.3)
BET	3.4 (1.1-5.6)	3.9 (0.8-8.8)
CIT	2.2 (1.2-4.0)	3.8 (2.2-9.7)
CTN	10.0 (1.7-24.6)	9.6 (1.1-43.4)
CNN	5.1 (2.5-8.1)	17.0 (6.6-63.4)
CYS	23.2 (3.8-39.4)	19.1 (6.0-46.8)
DMG	0.7 (0.2-1.9)	1.6 (0.4-4.0)
GSH	1.9 (0.5-4.7)	1.9 (0.4-6.7)
GLY	2.4 (1.7-5.7)	4.3 (1.8-11.7)
HARG	0.495 (0.192-0.777)	0.392 (0.104-1.7)
HCYS	0.830 (0.075-1.5)	0.734 (0.056-2.6)
MET	3.4 (1.4-8.2)	2.8 (1.4-5.8)
SAH	0.008 (0.001-0.035)	0.036 (0.006-0.319)
SAM	0.022 (0.003-0.043)	0.065 (0.040-0.157)
SDMA	0.075 (0.024-0.148)	0.350 (0.082-1.4)

4.10.2. UNIVARIATE ANALYSIS

Univariate statistics deals with observations of a single analyte in data, without taking into account its potential relation with other variables. With the aim of stating whether concentration difference between groups was significant or not, after testing for normality using Kolmogorov-Smirnov test, Student's *t*-test was applied for parametric variables, whereas *U* test in accordance to Mann and Whitney was applied to non-parametric variables using SPSS version 23 (IBM) software.

Despite the varying nature of CKD paediatrics involved in this study, significant differences ($p < 0.001$) were found in plasma concentrations of GLY, CIT, CNN, SDMA and ADMA when comparing paediatrics suffering from CKD and control patients, being *p*-value higher than 0.05 for the rest of the analytes.

It should be noted that plasma CNN concentration, used currently as a biomarker of renal function, differed quite a lot in paediatrics suffering from CKD due to differences in CKD stage as well as different renal replacement treatments received. Moreover, some of these CKD patients did not show increased CNN levels as a consequence of renal replacement therapy received. For that reason, with the aim of finding a relation between high concentrations of CNN and down-regulation or up-regulation of different metabolites, CKD paediatrics with CNN plasma concentrations higher than 12 µg/mL¹⁵⁹ were considered (a

total of 21 patients) and compared with control population. As a result, significant differences were found for DMG in addition to all of the previously reported analytes ($p < 0.001$ for all of them).

Table 4.14. Analytes showing significant differences between control and CKD. † Considering only CKD patients with increased plasma creatinine concentrations ($>12 \mu\text{g/mL}$), differences in DMG concentration were found ($p < 0.001$).

COMPOUND	Control	CKD	<i>p</i> -value
ADMA	0.133 (0.046-0.217)	0.180 (0.117-0.275)	$p < 0.001$
CIT	2.2 (1.2-4.0)	3.8 (2.2-9.7)	$p < 0.001$
CNN	5.1 (2.5-8.1)	17.0 (6.6-63.4)	$p < 0.001$
DMG	0.7 (0.2-1.9)	1.6 (0.4-4.0)	$p < 0.001^\dagger$
GLY	2.4 (1.7-5.7)	4.3 (1.8-11.7)	$p < 0.001$
SDMA	0.075 (0.024-0.148)	0.350 (0.082-1.4)	$p < 0.001$

Having a look at Table 4.13, SAH and SAM seem to be slightly increased in CKD population according to median values and IQR, despite this difference not being significant according to univariate analysis. For that reason, the possibility that SAM/SAH ratio, also called methylation index⁹⁸, could be affected was considered. SAM/SAH relation was calculated for the samples to verify whether methylation power might be affected in paediatrics suffering from CKD. However, no significant difference was found between both groups ($p > 0.05$). Thus, the use of a higher population might be suggested for future research to reach normal distribution of population and be able to clarify whether there might be any significant difference between both populations regarding SAM and SAH.

4.10.3. MULTIVARIATE ANALYSIS

4.10.3.1. Determination of outlying samples

Multivariate analysis approach is based on the observation and analysis of more than one statistical variable at a time. Due to the fact that outlying samples could affect multivariate models built, first of all, outlying samples were defined according to Hierarchical Clustering with single, complete and average linkage. Clustering deals with grouping samples according to object-wise similarities or distances and representing these distances in dendrograms. In general, single linkage clustering is more prone to provide individual groupings, whereas complete and average linkage clustering lead to more compact

clusters. Therefore, two big clusters could be expected for average and complete clustering methods: control and CKD groups. However, sample 11 clustered apart in all the clustering methods and diverse aspects of this sample were tracked since plasma treatment to data analysis. According to annotations taken prior to sample treatment, differences in plasma sample colour had been reported (a redish-pink colour) for this sample. Therefore, sample 11 was considered an outlying sample.

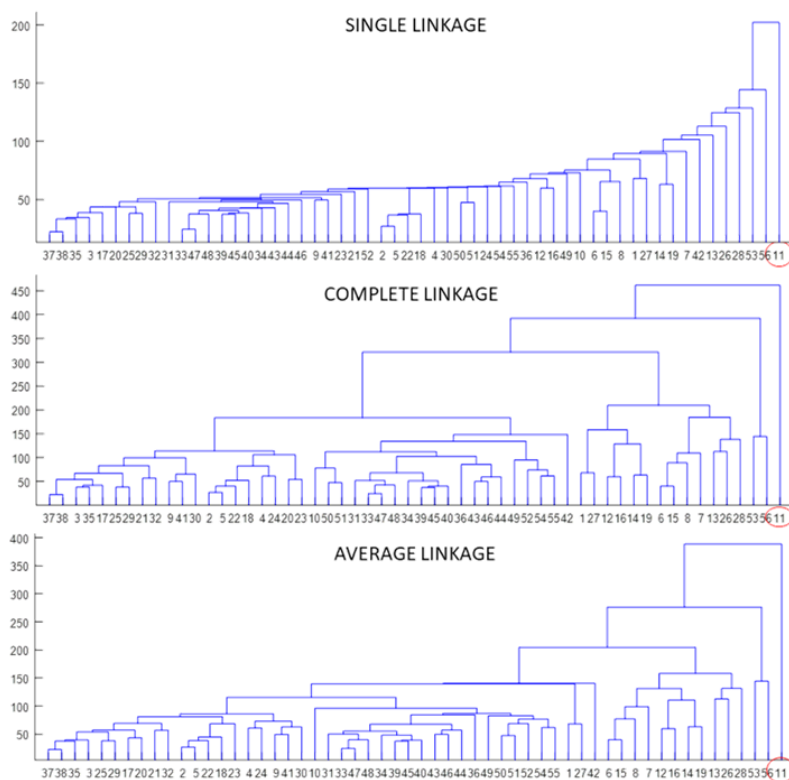


Fig. 4.39. Dendrograms obtained from different hierarchical clustering methods showing that sample 11 could correspond to an outlying sample.

4.10.3.2. Cross-correlations between metabolites

With the aim of evaluating possible relations between concentrations of different metabolites, cross-correlations of these analytes with CNN, current biomarker for CKD, were studied in the whole population. As observed in Table 4.15, the highest correlation was found for CNN-SDMA ($r=0.908$, $p<0.01$), CNN-CIT ($r=0.839$, $p<0.01$) and CNN-SAM ($r=0.773$, $p<0.05$). Regarding the interrelations between analytes that showed correlation with CNN, a high correlation was found between SDMA and SAM ($r=0.924$, $p<0.05$). In addition, CIT correlated well with SDMA ($r=0.837$, $p<0.01$) and SAM ($r=0.739$, $p<0.01$).

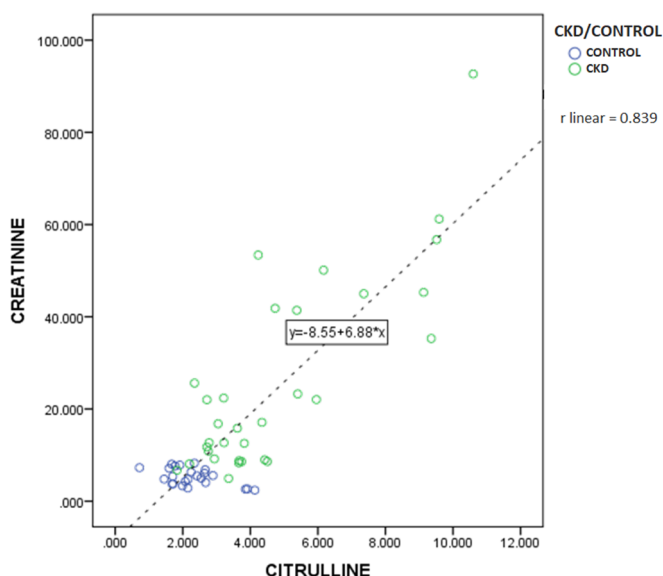


Fig. 4.40. Correlation between CNN and CIT concentrations ($\mu\text{g}/\text{mL}$) in the analyzed CKD and control plasma samples.

4.10.3.3. Scaling and construction of PCA model

Multivariate data analysis takes into consideration the intrinsic interdependency of metabolite concentrations. Regarding that concentrations of multiple analytes were non-normally distributed across samples, logarithm transformation was performed. Afterwards, autoscaling was carried out, i.e. the concentration of each metabolite in each sample was normalized by means of subtracting mean metabolite concentration and dividing it by their standard deviation. Autoscaling assumes that all metabolites are equally eligible to be important biomarkers, despite presenting higher or lower concentrations in plasma.

This log-transformed and autoscaled data set was subsequently subjected to PCA and the resulting biplot is represented in Fig. 4.41. The first three principal components accounted for 58 % of variance.

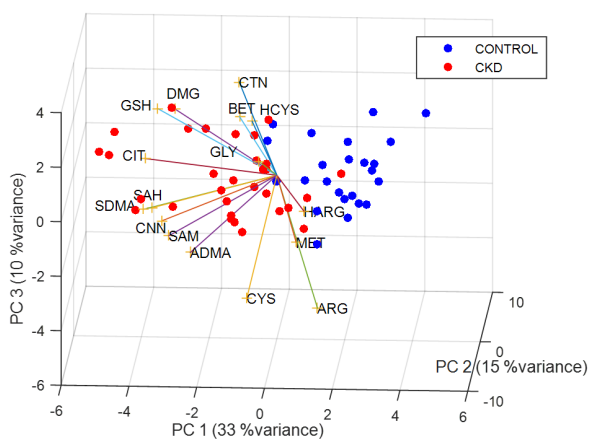


Fig. 4.41. PCA biplot containing the score plot for control and CKD samples and the loading plot with the 16 amino acids.

The PCA biplot obtained shows a good separation between control and CKD sample groups when colouring the samples according to their membership. In addition, the relation between metabolite concentrations and sample groups can be obtained from this representation. SDMA, SAH, CNN, SAM, CIT, ADMA, GSH, DMG and GLY were found to be increased in CKD patients comparing with control samples.

4.10.3.4. Classification between early CKD and control samples

As there is some interest in finding biomarkers to improve the early diagnosis of CKD, a PCA model was constructed and a classification method was selected considering only early disease samples (those at CKD2 stage) and control samples (see Fig. 4.42).

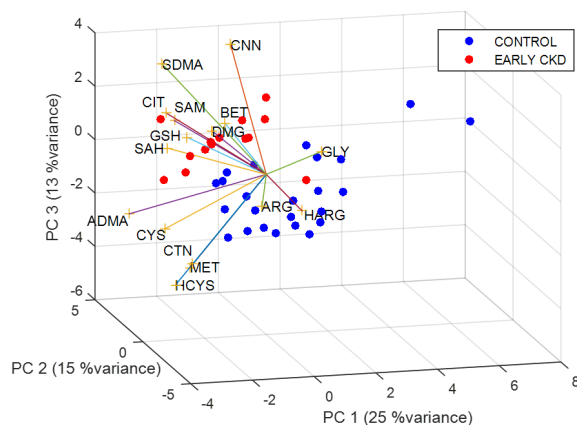


Fig. 4.42. PCA biplot containing control and CKD stage 2 samples with the 16 amino acids.

As it can be observed in Fig. 4.42, separation between groups was complete except for one CKD sample, which was placed with control samples. Regarding the relation between sample groups and metabolites, a similar distribution of the loadings was obtained for PCA containing all CKD samples and PCA containing only early CKD samples in addition to control samples.

Moreover, classification of the samples according to a model including only CNN and using all the metabolites were compared (Table 4.16). The best classification method for CNN was QDA with a performance of 71 %, whereas, after the optimization of the classifiers, using all the metabolites with PLS-DA with 2 LV the percentage of success was 76 % and with SPLS-DA with 1 LV and 4 metabolites was 67 %. CIT, CNN, SAM and SDMA were the output metabolites for SPLS-DA variable reduction and classification method. In addition, with the aim of having an idea of the weight of CNN on the model, performance was also calculated using all the metabolites except for CNN (15 analytes). PLS-DA implementation with 2 LV returned a performance of 59 %. Thus, it can be concluded that despite being CNN alone better than other 15 metabolites for prediction purposes, these metabolites together with CNN could be useful biomarkers.

Table 4.16. Summary of performance using optimized classification methods.

DATA	Optimal method	Performance
Only CNN	QDA	71 %
All metabolites	PLS-DA (LV=2)	76 %
	SPLS-DA (LV=1, nvar=4: CIT, CNN, SAM, SDMA)	67 %
All except for CNN	PLS-DA (LV=2)	59 %

The most practical approach and the most convenient for later clinical applicability relies on selecting the minimum number of metabolites capable of increasing prediction accuracy compared to CNN alone, and if possible also an increased one in comparison with using all the 16 metabolites. As showed previously, CNN, CIT, SAM and SDMA metabolites showed a similar efficiency for the early diagnosis of CKD as SPLS-DA method output in comparison with the prediction achieved with CNN alone. Therefore, it would be of special interest obtaining the classification performance of the model using only these four analytes in the input dataset.

PCA model was built including only these 4 metabolites (Fig. 4.43). As it can be observed in this figure, separation in the PCA model of CKD and control groups was almost complete and variance of the model increases to 92 % for 3 PC. In addition, the possibility of using

these metabolites as an input was evaluated with the aim of improving classification accuracy. Once again, the different classifiers were optimized for CNN, CIT, SAM and SDMA metabolites and a performance of 89 % was achieved using PLS-DA (LV=1) for the classification of early CKD and control samples. This means an increase in performance of 18 % when using CIT, SAM and SDMA in addition to CNN in comparison with using CNN only.

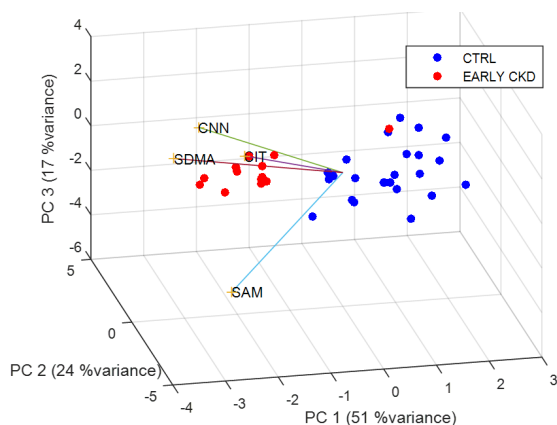


Fig. 4.43. PCA biplot containing control and early CKD samples with CIT, CNN, SDMA and SAM.

Regarding that CNN is the classical biomarker used for CKD diagnosis, the effect of using CIT, SDMA and SAM alone was also evaluated, in order to assess the contribution of these new metabolites (see Fig. 4.44).

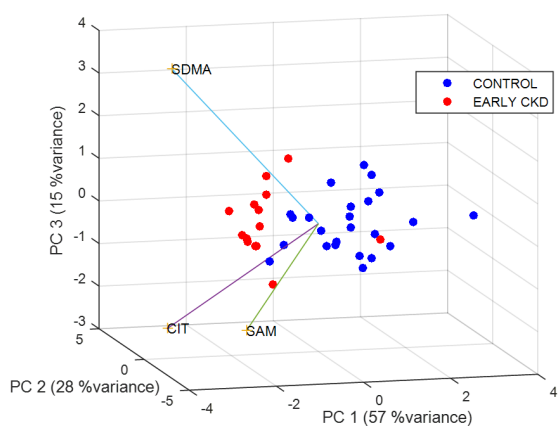


Fig. 4.44. PCA biplot containing control and early CKD samples with CIT, SDMA and SAM.

According to the PCA biplot showed in Fig. 4.44, CIT, SDMA and SAM alone enabled separation of early CKD and control groups, even though separation was slightly better using CIT, SDMA and SAM in addition to CNN. From the comparison of both PCA biplots it

can be concluded that CIT, SDMA and SAM metabolites could be contribute to CKD diagnosis usually performed using CNN alone.

4.10.3.5. Stage-independent CKD diagnosis

Even though the main goal of this work deals with finding a few potential biomarkers for early diagnosis, it would be interesting to check whether these 16 metabolites could be valuable regardless of the stage of CKD to be applicable in the clinical setting. For that reason, the same procedure was repeated using the 16 metabolites for all CKD samples, regardless of the CKD stage.

In this case, PLS-DA (LV=1) of the 16 analytes of interest resulted in a classification accuracy of 84 %. Using CNN alone to classify the samples as control or CKD gave a performance of 81 %, being QDA again the best classification method. From these results it can be inferred that, as in both cases classification performance was similar, using all metabolites would neither improve nor jeopardize the accuracy in the diagnosis.

Concerning that CIT, SAM, SDMA and CNN provided the best results for early CKD diagnosis, PCA and classification was repeated using as an input the concentration in plasma of these analytes in all patients, regardless of the CKD stage (Fig. 4.45). Once applied PLS-DA (LV=1) to the total amount of samples a performance of 91 % was achieved, thus enabling an improvement of 10 % in comparison with CNN alone. In addition, PCA was also obtained for all the metabolites except for CNN to evaluate the contribution of SAM, CIT and SDMA in group differentiation, which showed to be important (see Fig. 4.45).

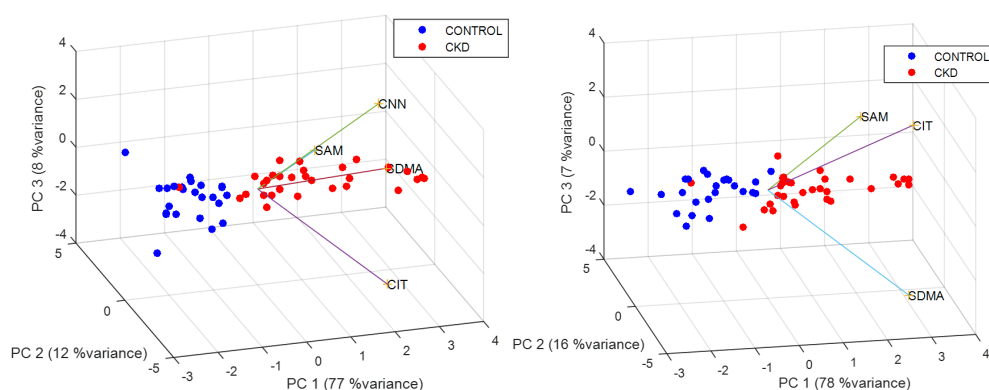


Fig. 4.45. PCA biplot containing control and all CKD samples with CIT, CNN, SDMA and SAM (left) and the same metabolites except for CNN (right).

From these results, it could be stated that metabolites CIT, SAM and SDMA improve the accuracy of CNN diagnosis not only for early disease samples but also when using all the samples, regardless of the CKD stage. Accordingly, the use of CIT, SAM and SDMA might be of special interest to add to CNN, as they improve the accuracy of CNN diagnosis alone.

4.10.3.6. Classification performance for CKD stages

Different equations have been developed over the years with the aim of diagnosing those patients suffering from CKD and classifying them into different stages according to the GFR calculated from the equation, such as: Schwartz¹⁶⁰, Bedside Schwartz equation¹⁶¹, Modification of Diet in Renal Disease (MDRD)^{162,163}, Counahan-Barratt¹⁶⁴ and Cockcroft-Gault¹⁶⁵.

“Gold standards” like iohexol, iothalamate, inuline, ⁵¹Cr-EDTA or ^{99m}Tc-DTPA are generally used to get real GFR value and adjust the equations to approximate estimated GFR value based on the measurement of endogenous metabolites. Despite this procedure being the ideal, metabolites were evaluated using PCA and different classification methods were used to determine the stage of the disease as a preliminary study prior to the development of any new equation.

PCA model was built both for the 16 metabolites as well as for the selected 4 metabolites, and coloured according to the different CKD stages.

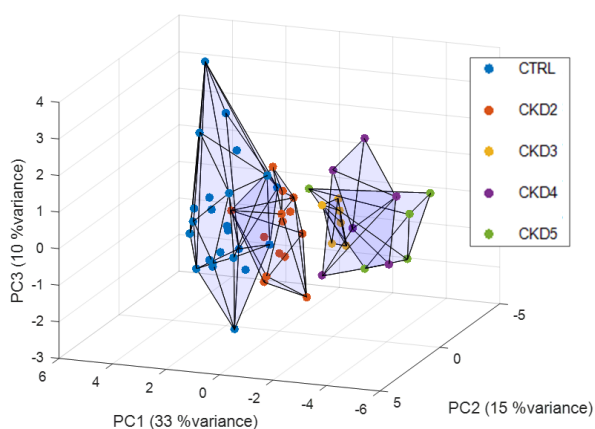


Fig. 4.46. Convex hull representation of PCA using the 16 metabolites.

PCA model obtained using all the metabolites (see Fig. 4.46) showed that complete separation was achieved between the groups made of CTRL-CKD2 and CKD3 to CKD5

groups. In addition, taking into account the loadings, concentrations of some metabolites increased with disease severity, such as SAH, SAM, SDMA, ADMA, CNN, CIT, GSH, GLY and DMG.

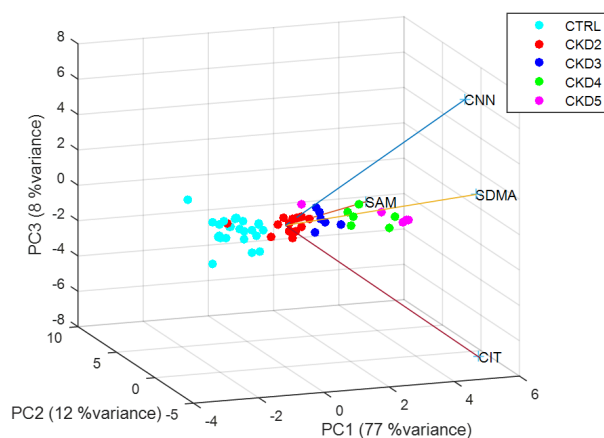


Fig. 4.47. PCA biplot showing the separation of control and different CKD stage patients.

Likewise, as showed in the PCA biplot from Fig.4.47, CIT, SAM, SDMA and CNN allowed a good separation of the groups and a gradation in accordance with the severity of CKD. Moreover, even in the case in which CNN was not included, clear groups and a gradation could be observed in samples according to CKD stage (see Fig. 4.48).

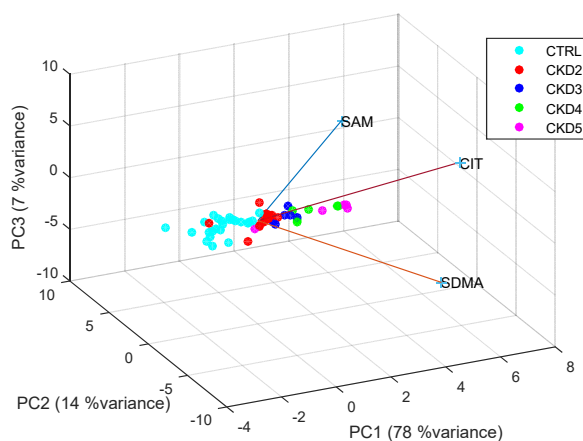


Fig. 4.48. PCA showing control and CKD sample distribution according to SAM, CIT and SDMA plasma concentrations.

These results show that the inclusion of SAM, CIT and SDMA in addition to CNN in equations for GFR determination and estimation of CKD might be promising.

Besides, to check whether these metabolites add, the accuracy of classification using the plasma concentrations of the proposed metabolites was compared with the use of CNN

concentration only. When using all the 16 metabolites (including CNN) with PLS-DA and 2 LV (66 %) some increase in performance was found and this increase was even higher using CIT, SAM, SDMA and CNN instead with PLS-DA and 3 LV (74 %), in comparison with using only CNN with QDA classification (52 %).

Despite the fact that these accuracies do not seem to be high, closer inspection of the confusion matrices shows that the majority of the misclassifications occur one level above or below in comparison with the CKD stage assessed by the nephrologist. Moreover, less than 3 % of the samples are misclassified 2 stages or more above or below. An average confusion matrix of classifications obtained with CIT, SAM, SDMA and CNN concentrations is showed in Table 4.17.

Table 4.17. Confusion matrix showing an average classification obtained after repeating 50 different divisions of the samples, building PLS-DA model with CIT, SAM, SDMA and CNN concentration, and calculating the classification for test samples.

	CONTROL pred	CKD2 pred	CKD3 pred	CKD4 pred	CKD5 pred
CONTROL known	3.1	0.4	0.0	0.0	0.0
CKD2 known	0.6	1.9	0.2	0.0	0.0
CKD3 known	0.1	0.5	1.1	0.1	0.2
CKD4 known	0.0	0.0	0.2	1.1	0.3
CKD5 Known	0.0	0.0	0.0	0.3	0.9

In the clinical setting, misclassification of the samples into different stages would affect the selected treatment for the patient. Nevertheless, the action plans recommended for each stage are general and do not change substantially from one stage to the level below or above. Therefore, these results are positive towards further study of SDMA, CNN, SAM and CIT as new biomarkers of CKD.

It has to be noted that classification performance for CNN alone might appear low in our predictions due to the fact that age, sex, height and/or weight are usually included in different equations to correct for some factor affecting serum CNN concentration. For instance, Schwartz equation uses height to correct for CNN production, as it is related with body mass and among the variables of body size tested by them, body length divided by CNN concentration provided the best correlation with GFR¹⁶⁰.

4.10.3.7. Age-, sex- and renal replacement therapy-related effects

The repercussion of different factors on metabolite concentrations, such as sex, age or treatment, need to be studied, in case any of these effects are present to be able to correct them in future equations.

With that aim, first of all, the effect of **age** was studied. PCA using CIT, CNN, SAM and SDMA as input variables was applied on the samples using two age groups ranging from 2 to 12 years old and from 13 to 18 years old. These groups of ages were selected in accordance with the division made by Way et al., according to the normal GFR values in paediatrics¹⁶⁶.

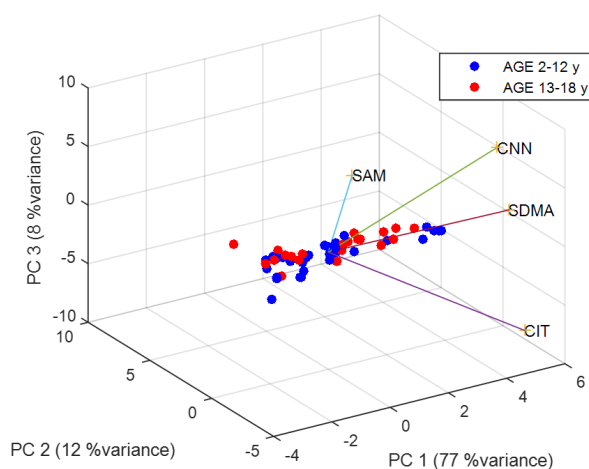


Fig. 4.49. PCA displaying children and adolescent patient distribution according to SAM, CNN, CIT and SDMA plasma concentrations.

As it can be observed in Fig. 4.49, although there is not any clear separation between children and adolescents, some trend could be observed. For that reason, with the aim of finding any relation between metabolite levels and age, Partial Least Squares (PLS) regression method was carried out to predict age according to plasma metabolite concentration (see Fig. 4.50). This regression was performed taking into account all samples together as well as CKD and control samples separately.

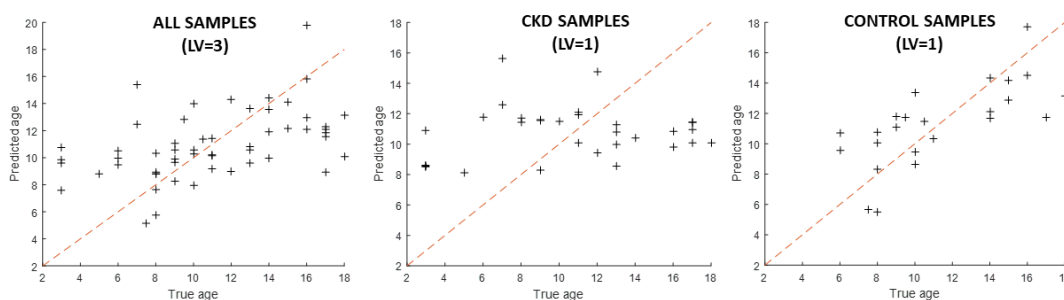


Fig. 4.50. PLS regression with optimized LV for all samples and for CKD and control samples separately.

When considering all samples or only CKD samples no good prediction was obtained, whereas when taking into account only control samples the prediction improved. Thus, it could be stated that there is some positive correlation between age and metabolite concentration in healthy paediatrics. These predictions were similar either using all the metabolites or using only CNN, SDMA, SAM and CIT as input variables.

In addition, the effect of **sex**, was studied by means of another PCA (Fig. 4.51). In this case, samples from both male and female groups were overlapped, being concluded that there is not any difference in metabolite concentrations regarding sex. This is not unexpected at all, since the patients were paediatrics, and minor differences concerning sex are presumed at these ages.

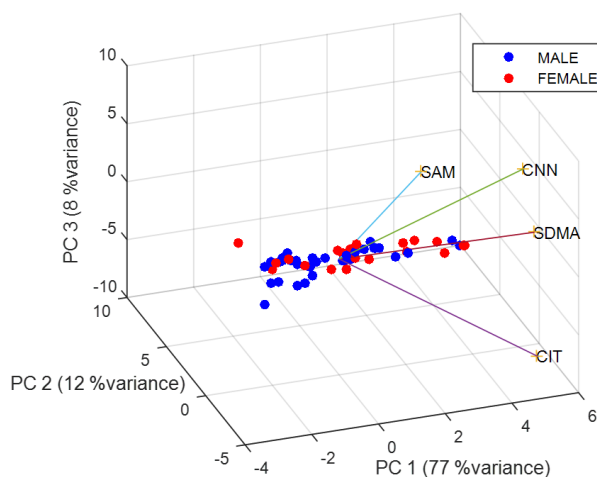


Fig. 4.51. PCA biplot coloured by sex using CIT, CNN, SDMA and SAM concentrations.

Besides, the distribution of males and females was also studied considering only patients aged 13-18 years old. However, sample distribution showed a similar pattern and no clear separation was observed neither using all nor the 4 selected metabolites.

Finally, the relation between metabolite concentration and **renal replacement therapy** (RRT) received by CKD patients was assayed. PCA was performed for CKD patients transplanted, dialyzed and not receiving RRT (see Fig. 4.52).

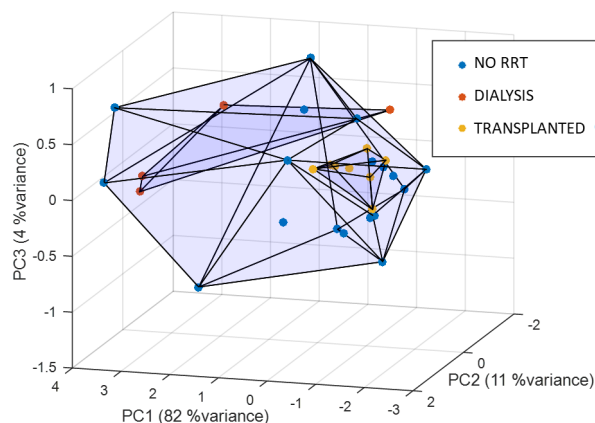


Fig. 4.52. Convex hull showing distribution of CKD patients according to the RRT received.

According to the obtained PCA representation, there is no separation between groups. However, the amino acid profile from transplanted patients showed to be more homogeneous in comparison with patients not receiving RRT and dialyzed patients. This could mean that amino acid profile tends to equalize after receiving a transplant. Indeed, even if control and transplanted groups keep separated in PCA, the distribution of the samples is similar, as showed in Fig. 4.53. This means that there are still some differences between both amino acid profiles.

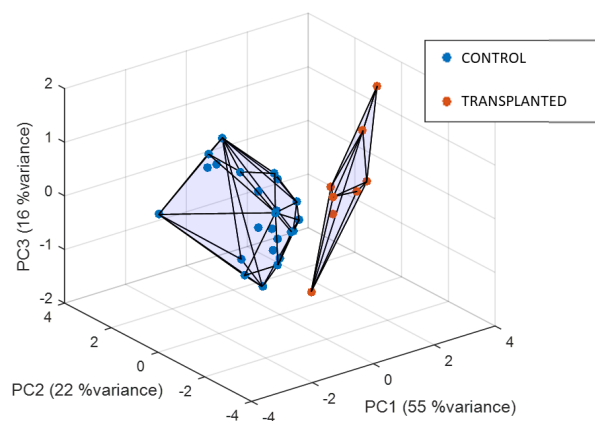


Fig. 4.53. Convex hull representation showing similar scattering of control and transplanted patients in PCA.

Patients eligible for the transplant were those for which other renal replacement therapy had failed. This homogenization in the metabolite profile after transplant could be expected

due to the fact that samples were collected at least a year after having received the transplant.

4.11. BIOLOGICAL INTERPRETATION OF THE RESULTS

The information obtained from data analysis was compared with the literature to verify if these up- or down-regulations could be explained by any known mechanism in arginine-creatinine pathway, arginine methylation or urea cycle.

Different biological interpretations were found for increased plasma levels of CIT, SDMA and SAM, the potential metabolites found to be especially interesting for CKD and control group differentiation, in addition to the classical biomarker CNN.

CIT increase could be explained by low activities of arginine/glycine amidinotransferase (AGAT) enzymes². In addition, SDMA isomer is mainly excreted in urine and thus, plasma SDMA rise could be a consequence of poor excretion due to renal impairment¹⁶⁷. Regarding SAM, its plasma levels might be increased because of renal injury. Indeed, kidney play a major role in aminothiol metabolism, and SAM levels depend largely on HCYS levels. Even though, it has to be noted that despite an association has been found between GFR and SAM, literature shows that HCYS may not correlated with GFR¹⁶⁸.

According to the univariate data analysis carried out, GLY, CIT, CNN, SDMA, ADMA and DMG would be increased in CKD patients. The explanation for GLY rise could be the same as for CIT, as both take part in a biochemical reaction regularized by AGAT enzyme, which could be affected in renal patients². Concerning ADMA, 80 % is eliminated through metabolization by means of the enzyme dimethylarginine dimethylaminohydrolase (DDAH) and 20 % by renal excretion, and impaired excretion may cause ADMA increase¹⁶⁷. Additionally, an exaggerated endogenous synthesis of ADMA has also been suggested, which could act as a contributing factor for CKD progression¹⁶⁷. Regarding DMG, this feedback inhibitor of betaine-homocysteine methyltransferase enzyme is normally excreted in urine or metabolized to sarcosine²⁶. Thus, it is suggested that it might be accumulated due to impaired urinary excretion in addition to any unknown factor that may also be involved.

Finally, from multivariate analysis SAH and GSH have also been found to be affected in paediatrics with CKD, in addition to previously described increases of SDMA, GLY, CNN,

CIT, ADMA, SAM and DMG. Regarding SAH, as explained before, kidney is involved in aminothiols metabolism and similarly to SAM, SAH levels also depend largely on HCYS levels, which might be affected if kidney is impaired. Moreover, a reduction in metabolization and/or excretion ability of SAH has been suggested to be a primary event in CKD¹⁶⁹. Several biomarkers of oxidative stress like GSH might be increased in patients suffering from CKD¹⁷⁰. The nature of the oxidative stress in CKD patients could be related to increased reactive oxygen species (ROS) production and reduced clearance, or even a consequence of ineffective antioxidant defense mechanism. Nevertheless, it should be noted that even if some previous studies had showed increased total GSH levels in adults suffering from CKD¹⁷¹, some other stated the opposite¹⁷² or did not find any significant difference between healthy and CKD patients⁶⁸.

4.12. CONCLUSIONS

Taking into consideration the proposed objectives and the results obtained in this targeted metabolomics study, it can be concluded that:

- ✓ A new LC-QTOF-MS analytical methodology has been optimized for the analysis of amino acids, amino acid derivatives, and related compounds from the arginine-creatine pathway, the arginine methylation, and the urea cycle which could be of interest for the diagnosis of CKD in paediatric patients. This new method enables the identification and quantification of 16 compounds in a single run in MS and MS/MS mode in less than 19 min.
- ✓ Plasma sample treatment requires a low sample volume and involves the addition of dithiothreitol to obtain total aminothiol concentration, which avoids under- or overestimation in the quantification of these compounds. This step is followed by a simple plasma protein precipitation process and avoids the need for derivatization.
- ✓ After optimization and validation, this method was successfully applied to the analysis of both a heterogeneous group of CKD paediatric patients and a control population, being significant differences found for GLY, CIT, CNN, ADMA, SDMA and DMG according to univariate data analysis.
- ✓ Three new metabolites, CIT, SAM and SDMA, have been proposed as potential biomarkers in addition to the commonly used CNN to be implemented since they enable a better diagnosis of early stages of CKD in paediatrics. Besides, CIT, SAM, SDMA and

CNN showed greater improvement of the diagnosis of all the stages of CKD, and some gradation in concentration was found in accordance with the CKD stage.

- ✓ It is reasonable to think of including these 3 metabolites in a new equation to be used in the screening tests carried out by general practitioners to approximate the prediction of the disease made by nephrologists and foresee the stage of CKD, before referring them to the specialist.
- ✓ The collection of a higher number of samples is suggested to obtain GFR by means of some contrast agent like iohexol, used as "gold standard", to adjust a new equation for GFR calculation including CIT, SDMA and SAM and compare the performances obtained from this equation with other commonly used equations in paediatrics like Schwartz.
- ✓ This work resulted in the following publications in peer-reviewed scientific journals, which are annexed at the end of this manuscript:

"Benito S, Sánchez A, Unceta N, Andrade F, Aldámiz-Echevarria L, Goicolea MA, Barrio RJ. *LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease*. *Anal Bioanal Chem*. 2016; 408(3):747-60. doi: 10.1007/s00216-015-9153-9."

"Benito S, Sanchez-Ortega A, Unceta N, Jansen JJ, Postma G, Andrade F, Aldámiz-Echevarria L, Buydens LMC, Goicolea MA, Barrio RJ. *Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data*. *J Pharm Biomed Anal*. [In press] doi: 10.1016/j.jpba.2017.10.036. "

4.13. BIBLIOGRAPHY

1. Vanholder R, Argiles A, Baurmeister U, Brunet P, Clark W, Cohen G, et al. Uremic toxicity: present state of the art. *Int J Artif Organs*. 2001; 24: 695-725.
2. Andrade F, Rodriguez-Soriano J, Prieto JA, Elorz J, Aguirre M, Ariceta G, et al. The Arginine-Creatine Pathway is Disturbed in Children and Adolescents With Renal Transplants. *Pediatr Res*. 2008; 64: 218-222.
3. Duranton F, Cohen G, De Smet R, Rodriguez M, Jankowski J, Vanholder R, et al. Normal and pathologic concentrations of uremic toxins. *J Am Soc Nephrol*. 2012; 23: 1258-1270.
4. Wu Q, Lai X, Zhu Z, Hong Z, Dong X, Wang T, et al. Evidence for Chronic Kidney Disease-Mineral and Bone Disorder Associated With Metabolic Pathway Changes. *Medicine*. 2015; 94: e1273.
5. Cibulka R, Racek J. Metabolic disorders in patients with chronic kidney failure. *Physiol Res*. 2007; 56: 697-705.
6. Van de Poll MCG, Soeters PB, Deutz NEP, Fearon KCH, Dejong CHC. Renal metabolism of amino acids: Its role in interorgan amino acid exchange. *Am J Clin Nutr*. 2004; 79: 185-197.
7. Kyoto Encyclopedia of Genes and Genomes (KEGG). Ko00220 - Arginine biosynthesis [last accessed: 15/11/2017]. Available from: http://www.genome.jp/kegg-bin/show_pathway?ko00220.
8. Andrade F, Rodriguez-Soriano J, Prieto JA, Aguirre M, Ariceta G, Lage S, et al. Methylation cycle, arginine-creatine pathway and asymmetric dimethylarginine in pediatric renal transplant. *Nephrol Dial Transpl*. 2011; 26: 328-336.
9. Kyoto Encyclopedia of Genes and Genomes (KEGG). Ko00020 - Citrate cycle (TCA) [last accessed: 15/11/2017]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map00020.
10. Kyoto Encyclopedia of Genes and Genomes (KEGG). Hsa00330- Arginine and proline metabolism [last accessed: 15/11/2017]. Available from: http://www.genome.jp/kegg-bin/show_pathway?hsa00330.
11. Kyoto Encyclopedia of Genes and Genomes (KEGG). Map01230 - Biosynthesis of amino acids [last accessed: 15/11/2017]. Available from: http://www.genome.jp/dbget-bin/www_bget?map01230.
12. Kyoto Encyclopedia of Genes and Genomes (KEGG). Map01200 - Carbon metabolism [last accessed: 15/11/2017]. Available from: http://www.genome.jp/kegg-bin/show_pathway?map=map01200.
13. Michal G, Schomburg D. Introduction and General Aspects. In: Michal G, Schomburg D, editors. *Biochemical Pathways: an atlas of biochemistry and molecular biology*. 2nd ed: John Wiley & Sons, Inc; 2012; p. 1-13.

14. Erlandsen EJ, Randers E, Kristensen JH. Evaluation of the Dade Behring N Latex Cystatin C assay on the Dade Behring Nephelometer II system. *Scand J Clin Lab Invest*. 1999; 59: 1-8.
15. Nelson DL, Cox MM. Amino Acid Oxidation and the Production of Urea. In: Ahr K, editor. *Lehninger Principles of Biochemistry*. 5 ed: W.H. Freeman and company; 2008; p. 673-706.
16. Bedford MT, Richard S. Arginine methylation: An emerging regulator of protein function. *Mol Cell*. 2005; 18: 263-272.
17. Item CB, Stockler-Ipsiroglu S, Stromberger C, Muhl A, Alessandri MG, Bianchi MC, et al. Arginine:glycine amidinotransferase deficiency: the third inborn error of creatine metabolism in humans. *Am J Hum Genet*. 2001; 69: 1127-1133.
18. Fliser D, Kronenberg F, Kielstein JT, Morath C, Bode-Boeger SM, Haller H, et al. Asymmetric dimethylarginine and progression of chronic kidney disease: the mild to moderate kidney disease study. *J Am Soc Nephrol*. 2005; 16: 2456-2461.
19. Fleck C, Schweitzer F, Karge E, Busch M, Stein G. Serum concentrations of asymmetric (ADMA) and symmetric (SDMA) dimethylarginine in patients with chronic kidney diseases. *Clin Chim Acta*. 2003; 336: 1-12.
20. Marescau B, Nagels G, Possemiers I, De Broe ME, Becaus I, Billiouw J-M, et al. Guanidino compounds in serum and urine of nondialyzed patients with chronic renal insufficiency. *Metab Clin Exp*. 1997; 46: 1024-1031.
21. Chen GF, Baylis C. In vivo renal arginine release is impaired throughout development of chronic kidney disease. *Am J Physiol*. 2010; 298: F95-F102.
22. Mutsaers HAM, Engelke UFH, Wilmer MJG, Wetzels JFM, Wevers RA, Van den Heuvel LP, et al. Optimized metabolomic approach to identify uremic solutes in plasma of stage 3-4 chronic kidney disease patients. *PLoS One*. 2013; 8: e71199.
23. Lever M, Sizeland PCB, Bason LM, Hayman CM, Robson RA, Chambers ST. Abnormal glycine betaine content of the blood and urine of diabetic and renal patients. *Clin Chim Acta*. 1994; 230: 69-79.
24. De Deyn P, Marescau B, Lornoy W, Becaus I, Lowenthal A. Guanidino compounds in uraemic dialysed patients. *Clin Chim Acta*. 1986; 157: 143-150.
25. Suliman ME, Anderstam B, Lindholm B, Bergstrom J. Total, free, and protein-bound sulphur amino acids in uraemic patients. *Nephrol Dial Transpl*. 1997; 12: 2332-2338.
26. McGregor DO, Dellow WJ, Lever M, George PM, Robson RA, Chambers ST. Dimethylglycine accumulates in uremia and predicts elevated plasma homocysteine concentrations. *Kidney Int*. 2001; 59: 2267-2272.
27. Zwolinska D, Grzeszczak W, Szczepanska M, Makulska I, Kilis-Pstrusinska K, Szprynger K. Oxidative stress in children on peritoneal dialysis. *Periton Dialysis Int*. 2009; 29: 171-177.

28. Chuang CK, Lin SP, Chen HH, Chen YC, Wang TJ, Shieh WH, et al. Plasma free amino acids and their metabolites in Taiwanese patients on hemodialysis and continuous ambulatory peritoneal dialysis. *Clin Chim Acta*. 2006; 364: 209-216.
29. Holecek M, Sprongl L, Tilser I, Tichy M. Leucine and protein metabolism in rats with chronic renal insufficiency. *Exp Toxicol Pathol*. 2001; 53: 71-76.
30. Drechsler C, Kollerits B, Meinitzer A, Maerz W, Ritz E, Koenig P, et al. Homoarginine and progression of chronic kidney disease: results from the mild to moderate kidney disease study. *PLoS One*. 2013; 8: e63560.
31. Ravani P, Maas R, Malberti F, Pecchini P, Mieth M, Quinn R, et al. Homoarginine and mortality in pre-dialysis chronic kidney disease (CKD) patients. *PLoS One*. 2013; 8: e72694.
32. Merouani A, Lambert M, Delvin EE, Genest J, Jr., Robitaille P, Rozen R. Plasma homocysteine concentration in children with chronic renal failure. *Pediatr Nephrol*. 2001; 16: 805-811.
33. El Sawy MA, Zaki MM, El-Hakim IZ, Mowafy ME, Al-Abd HS. Serum amino acid abnormalities in pediatric patients with chronic renal failure with and without history of thromboembolic manifestations. *Egypt J Med Hum Genet*. 2012; 13: 73-80.
34. Loehrer FMT, Angst CP, Brummer FP, Haefeli WE, Fowler B. Evidence for disturbed S-adenosylmethionine:S-adenosylhomocysteine ratio in patients with end-stage renal failure: a cause for disturbed methylation reactions? *Nephrol Dial Transpl*. 1998; 13: 656-661.
35. Perna AF, Ingrosso D, Galletti P, Zappa V, De Santo NG. Membrane protein damage and methylation reactions in chronic renal failure. *Kidney Int*. 1996; 50: 358-366.
36. Dudley E, Yousef M, Wang Y, Griffiths WJ. Targeted metabolomics and mass spectrometry. *Adv Protein Chem Struct Biol*. 2010; 80: 45-83.
37. Patti GJ, Yanes O, Siuzdak G. Innovation Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*. 2012; 13: 263-269.
38. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2012; 883-884: 68-75.
39. Fiehn O. Metabolomics - the link between genotypes and phenotypes. *Plant Mol Biol*. 2002; 48: 155-171.
40. Zhou B, Xiao JF, Tuli L, Resson HW. LC-MS-based metabolomics. *Mol Biosyst*. 2012; 8: 470-481.
41. Ceglarek U, Kortz L, Leichtle A, Fiedler GM, Kratzsch J, Thiery J. Rapid quantification of steroid patterns in human serum by on-line solid phase extraction combined with liquid chromatography-triple quadrupole linear ion trap mass spectrometry. *Clin Chim Acta*. 2009; 401: 114-118.
42. Van Dyk M, Mangoni AA, McEvoy M, Attia JR, Sorich MJ, Rowland A. Targeted arginine metabolomics: A rapid, simple UPLC-QToF-MS^E based approach for

- assessing the involvement of arginine metabolism in human disease. *Clin Chim Acta*. 2015; 447: 59-65.
43. Krumpochova P, Bruyneel B, Molenaar D, Koukou A, Wuhrer M, Niessen WMA, et al. Amino acid analysis using chromatography-mass spectrometry: An inter platform comparison study. *J Pharm Biomed Anal*. 2015; 114: 398-407.
 44. Dejaegher B, Vander Heyden Y. HILIC methods in pharmaceutical analysis. *J Sep Sci*. 2010; 33: 698-715.
 45. Hemstrom P, Irgum K. Hydrophilic interaction chromatography. *J Sep Sci*. 2006; 29: 1784-1821.
 46. Ikegami T, Tomomatsu K, Takubo H, Horie K, Tanaka N. Separation efficiencies in hydrophilic interaction chromatography. *J Chromatogr A*. 2008; 1184: 474-503.
 47. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens J-P, Morla A, et al. ESI-MS/MS analysis of underivatized amino acids: A new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionization mode. *Rapid Commun Mass Sp*. 2003; 17: 1297-1311.
 48. Callejon RM, Troncoso AM, Morales ML. Determination of amino acids in grape-derived products: A review. *Talanta*. 2010; 81: 1143-1152.
 49. Piraud M, Vianey-Saban C, Bourdin C, Acquaviva-Bourdain C, Boyer S, Elfakir C, et al. A new reversed-phase liquid chromatographic/tandem mass spectrometric method for analysis of underivatized amino acids: Evaluation for the diagnosis and the management of inherited disorders of amino acid metabolism. *Rapid Commun Mass Sp*. 2005; 19: 3287-3297.
 50. Lau T, Owen W, Yu YM, Noviski N, Lyons J, Zurakowski D, et al. Arginine, citrulline, and nitric oxide metabolism in end-stage renal disease patients. *J Clin Invest*. 2000; 105: 1217-1225.
 51. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). L-Arginine (HMDB00517). [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00517>.
 52. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Asymmetric dimethylarginine (HMDB01539) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB01539>.
 53. MacAllister RJ, Fickling SA, Whitley GSJ, Vallance P. Metabolism of methylarginines by human vasculature; implications for the regulation of nitric oxide synthesis. *Br J Pharmacol*. 1994; 112: 43-48.
 54. Leiper JM, Santa Maria J, Chubb A, MacAllister RJ, Charles IG, Whitley GSJ, et al. Identification of two human dimethylarginine dimethylaminohydrolases with distinct tissue distributions and homology with microbial arginine deiminases. *Biochem J*. 1999; 343: 209-214.
 55. Ogawa T, Kimoto M, Sasaoka K. Purification and properties of a new enzyme, NG,NG-dimethylarginine dimethylaminohydrolase, from rat kidney. *J Biol Chem*. 1989; 264: 10205-10209.

56. Nijveldt RJ, van Leeuwen PAM, van Guldener C, Stehouwer CDA, Rauwerda JA, Teerlink T. Net renal extraction of asymmetrical (ADMA) and symmetrical (SDMA) dimethylarginine in fasting humans. *Nephrol Dial Transpl.* 2002; 17: 1999-2002.
57. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Betaine (HMDB00043) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00043>.
58. Lever M, Slow S. The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism. *Clin Biochem.* 2010; 43: 732-744.
59. Moriyama T, Garcia-Perez A, Olson AD, Burg MB. Intracellular betaine substitutes for sorbitol in protecting renal medullary cells from hypertonicity. *Am J Physiol.* 1991; 260: F494-F497.
60. McGregor DO, Dellow WJ, Robson RA, Lever M, George PM, Chambers ST. Betaine supplementation decreases post-methionine hyperhomocysteinemia in chronic renal failure. *Kidney Int.* 2002; 61: 1040-1046.
61. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Creatine (HMDB00064) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00064>.
62. KEGG Enzyme Database. ENZYME: 1.14.13.39 [last accessed: 15/11/2017]. Available from: http://www.genome.jp/dbget-bin/www_bget?ec:1.14.13.39.
63. Edison EE, Brosnan ME, Meyer C, Brosnan JT. Creatine synthesis: production of guanidinoacetate by the rat and human kidney in vivo. *Am J Physiol.* 2007; 293: F1799-F1804.
64. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Citrulline (HMDB00904) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00904>.
65. Taegtmeyer H, Ingwall JS. Creatine-A Dispensable Metabolite? *Circ Res.* 2013; 112: 878-880.
66. Hunley TE, Kon V, Ichikawa I. Glomerular Circulation and Function. In: Avner ED, Harmon WE, Niaudet P, Yoshikawa N, editors. *Pediatric Nephrology*. 6th ed, 2009; p. 46-50.
67. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Creatinine (HMDB00562) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00562>.
68. Himmelfarb J, McMenamin E, McMonagle E. Plasma aminothiols oxidation in chronic hemodialysis patients. *Kidney Int.* 2002; 61: 705-716.
69. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Cysteine (HMDB00574) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00574>.
70. Paulsen CE, Carroll KS. Cysteine-mediated redox signaling: Chemistry, biology, and tools for discovery. *Chem Rev.* 2013; 113: 4633-4679.

71. Wu G, Fang YZ, Yang S, Lupton JR, Turner ND. Glutathione metabolism and its implications for health. *J Nutr.* 2004; 134: 489-492.
72. Meister A. Glutathione. In: Aria IM, Jakoby WB, Popper H, Schachter D, Shafritz DA, eds, editors. *The Liver: Biology and Pathobiology.* 2 ed. New York: Raven Press; 1988; p. 401-417.
73. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Dimethylglycine (HMDB00092) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00092>.
74. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Glycine (HMDB00123) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00123>.
75. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Glutathione (HMDB00125) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00125>.
76. Dou L, Jourde-Chiche N, Faure V, Cerini C, Berland Y, Dignat-George F, et al. The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells. *J Thromb Haemostasis.* 2007; 5: 1302-1308.
77. Denzoin Vulcano LA, Soraci AL, Tapia MO. Homeostasis del glutati3n. *Acta bioqu3m cl3n latinoam.* 2013; 47: 529-539.
78. Pastore A, Federici G, Bertini E, Piemonte F. Analysis of glutathione: implication in redox and detoxification. *Clin Chim Acta.* 2003; 333: 19-39.
79. Peters WHM, van Schaik A, Peters JH, van Goor H. Oxidised- and total non-protein bound glutathione and related thiols in gallbladder bile of patients with various gastrointestinal disorders. *BMC Gastroenterol.* 2007; 7: No pp. given.
80. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Homo-L-arginine (HMDB00670) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00670>.
81. Marz W, Meinitzer A, Drechsler C, Pilz S, Krane V, Kleber ME, et al. Homoarginine, cardiovascular risk, and mortality. *Circulation.* 2010; 122: 967-975.
82. Valtonen P, Laitinen T, Lyyra-Laitinen T, Raitakari OT, Juonala M, Viikari JSA, et al. Serum L-homoarginine concentration is elevated during normal pregnancy and is related to flow-mediated vasodilatation. *Circ J.* 2008; 72: 1879-1884.
83. Hrabak A, Bajor T, Temesi A. Comparison of substrate and inhibitor specificity of arginase and nitric oxide (NO) synthase for arginine analogs and related compounds in murine and rat macrophages. *Biochem Biophys Res Commun.* 1994; 198: 206-212.
84. Knowles RG, Palacios M, Palmer RMJ, Moncada S. Formation of nitric oxide from L-arginine in the central nervous system: a transduction mechanism for stimulation of the soluble guanylate cyclase. *P Natl Acad Sci USA.* 1989; 86: 5159-5162.
85. Refsum H, Ueland PM, Nygard O, Vollset SE. Homocysteine and cardiovascular disease. *Annu Rev Med.* 1998; 49: 31-62.

86. De Bree A, Verschuren WMM, Kromhout D, Kluijtmans LAJ, Blom HJ. Homocysteine determinants and the evidence to what extent homocysteine determines the risk of coronary heart disease. *Pharmacol Rev.* 2002; 54: 599-618.
87. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Homocysteine (HMDB00742) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00742>.
88. Jacobsen DW. Homocysteine and vitamins in cardiovascular disease. *Clin Chem.* 1998; 44: 1833-1843.
89. Massy ZA. Importance of homocysteine, lipoprotein (a) and non-classical cardiovascular risk factors (fibrinogen and advanced glycation end-products) for atherogenesis in uremic patients. *Nephrol Dial Transpl.* 2000; 15: 81-91.
90. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). L-Methionine (HMDB00696) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00696>.
91. Rafii M, Elango R, House JD, Courtney-Martin G, Darling P, Fisher L, et al. Measurement of homocysteine and related metabolites in human plasma and urine by liquid chromatography electrospray tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2009; 877: 3282-3291.
92. Garibotto G, Valli A, Anderstam B, Eriksson M, Suliman ME, Balbi M, et al. The kidney is the major site of S-adenosylhomocysteine disposal in humans. *Kidney Int.* 2009; 76: 293-296.
93. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). S-Adenosylhomocysteine (HMDB00939) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB00939>.
94. Kloor D, Stumvoll W, Schmid H, Kompf J, Mack A, Osswald H. Localization of S-adenosylhomocysteine hydrolase in the rat kidney. *J Histochem Cytochem.* 2000; 48: 211-218.
95. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). S-Adenosylmethionine (HMDB01185) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB01185>.
96. Finkelstein JD, Martin JJ. Methionine metabolism in mammals. Distribution of homocysteine between competing pathways. *J Biol Chem.* 1984; 259: 9508-9513.
97. Gellekink H, van Oppenraaij-Emmerzaal D, van Rooij A, Struys EA, den Heijer M, Blom HJ. Stable-isotope dilution liquid chromatography-electrospray injection tandem mass spectrometry method for fast, selective measurement of S-adenosylmethionine and S-adenosylhomocysteine in plasma. *Clin Chem.* 2005; 51: 1487-1492.
98. Stabler SP, Allen RH. Quantification of serum and urinary S-adenosylmethionine and S-adenosylhomocysteine by stable-isotope-dilution liquid chromatography-mass spectrometry. *Clin Chem.* 2004; 50: 365-372.

99. Human Metabolome Database (HMDB), The Metabolomics Innovation Centre (TMIC). Symmetric dimethylarginine (HMDB03334) [last accessed: 15/11/2017]. Available from: <http://www.hmdb.ca/metabolites/HMDB03334>.
100. Lee JH, Cook JR, Yang ZH, Mirochnitchenko O, Gunderson SI, Felix AM, et al. PRMT7, a new protein arginine methyltransferase that synthesizes symmetric dimethylarginine. *J Biol Chem*. 2005; 280: 3656-3664.
101. Richard S, Morel M, Cleroux P. Arginine methylation regulates IL-2 gene expression: a role for protein arginine methyltransferase 5 (PRMT5). *Biochem J*. 2005; 388: 379-386.
102. Bedford MT. Arginine methylation at a glance. *J Cell Sci*. 2007; 120: 4243-4246.
103. Kaspar H, Dettmer K, Gronwald W, Oefner PJ. Automated GC-MS analysis of free amino acids in biological fluids. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2008; 870: 222-232.
104. Rougé C, Des Robert C, Robins A, Le Bacquer O, De La Cochetiere MF, Darmaun D. Determination of citrulline in human plasma, red blood cells and urine by electron impact (EI) ionization gas chromatography-mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2008; 865: 40-47.
105. Beckmann B, Gutzki FM, Tsikas D. Sensitivity enhancement of a GC-MS/MS method for asymmetric dimethylarginine (ADMA) by plasma ultrafiltrate reduction. *Anal Biochem*. 2008; 372: 264-266.
106. Husek P, Simek P, Hartvich P, Zahradnickova H. Fluoroalkyl chloroformates in treating amino acids for gas chromatographic analysis. *J Chromatogr A*. 2008; 1186: 391-400.
107. Armstrong M, Jonscher K, Reisdorph NA. Analysis of 25 underivatized amino acids in human plasma using ion-pairing reversed-phase liquid chromatography/time-of-flight mass spectrometry. *Rapid Commun Mass Sp*. 2007; 21: 2717-2726.
108. Qu J, Wang Y, Luo G, Wu Z, Yang C. Validated quantitation of underivatized amino acids in human blood samples by volatile ion-pair reversed-phase liquid chromatography coupled to isotope dilution tandem mass spectrometry. *Anal Chem*. 2002; 74: 2034-2040.
109. Dietzen DJ, Weindel AL, Carayannopoulos MO, Landt M, Normansell ET, Reimschisel TE, et al. Rapid comprehensive amino acid analysis by liquid chromatography/tandem mass spectrometry: comparison to cation exchange with post-column ninhydrin detection. *Rapid Commun Mass Sp*. 2008; 22: 3481-3488.
110. Loscos N, Segurel M, Dagan L, Sommerer N, Marlin T, Baumes R. Identification of S-methylmethionine in Petit Manseng grapes as dimethyl sulphide precursor in wine. *Anal Chim Acta*. 2008; 621: 24-29.
111. Cao ZY, Sun LH, Mou RX, Zhou R, Zhu ZW, Chen MX. A novel method for the simultaneous analysis of seven biothiols in rice (*Oryza sativa* L.) using hydrophilic interaction chromatography coupled with electrospray tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2015; 976-977: 19-26.

112. Ngo S, Li X, O'Neill R, Bhoothpur C, Gluckman P, Sheppard A. Elevated S-adenosylhomocysteine alters adipocyte functionality with corresponding changes in gene expression and associated epigenetic marks. *Diabetes*. 2014; 63: 2273-2283.
113. Martens-Lobenhoffer J, Bode-Boeger SM. Mass spectrometric quantification of L-arginine and its pathway related substances in biofluids: The road to maturity. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2013; 964: 89-102.
114. Spagou K, Wilson ID, Masson P, Theodoridis G, Raikos N, Coen M, et al. HILIC-UPLC-MS for Exploratory Urinary Metabolic Profiling in Toxicological Studies. *Anal Chem*. 2011; 83: 382-390.
115. Oosterink JE, Buijs N, van Goudoever JB, Schierbeek H. A novel method for simultaneous measurement of concentration and enrichment of NO synthesis-specific amino acids in human plasma using stable isotopes and LC/MS ion trap analysis. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2014; 958: 10-15.
116. Harder U, Koletzko B, Peissner W. Quantification of 22 plasma amino acids combining derivatization and ion-pair LC-MS/MS. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2011; 879: 495-504.
117. Leung EMK, Wan C. A novel reversed-phase HPLC method for the determination of urinary creatinine by pre-column derivatization with ethyl chloroformate: comparative studies with the standard Jaffe and isotope-dilution mass spectrometric assays. *Anal Bioanal Chem*. 2014; 406: 1807-1812.
118. Yoon HR. Determination of plasma dibasic amino acids following trimethylsilyl-trifluoroacetyl derivatization using gas chromatography-mass spectrometry. *Arch Pharmacol Res*. 2013; 36: 366-373.
119. Martens-Lobenhoffer J, Bode-Boeger SM. Quantification of L-arginine, asymmetric dimethylarginine and symmetric dimethylarginine in human plasma: A step improvement in precision by stable isotope dilution mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2012; 904: 140-143.
120. Martens-Lobenhoffer J, Sulyok E, Czeiter E, Bueki A, Kohl J, Firsching R, et al. Determination of cerebrospinal fluid concentrations of arginine and dimethylarginines in patients with subarachnoid hemorrhage. *J Neurosci Methods*. 2007; 164: 155-160.
121. Tang YB, Teng L, Sun F, Wang XL, Peng L, Cui YY, et al. Determination of glycine in biofluid by hydrophilic interaction chromatography coupled with tandem mass spectrometry and its application to the quantification of glycine released by embryonal carcinoma stem cells. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2012; 905: 61-66.
122. Tsikas D, Beckmann B, Gutzki FM, Jordan J. Simultaneous gas chromatography-tandem mass spectrometry quantification of symmetric and asymmetric dimethylarginine in human urine. *Anal Biochem*. 2011; 413: 60-62.
123. Atzler D, Mieth M, Maas R, Boeger RH, Schwedhelm E. Stable isotope dilution assay for liquid chromatography-tandem mass spectrometric determination of L-homoarginine in human plasma. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2011; 879: 2294-2298.

124. Brown CM, Becker JO, Wise PM, Hoofnagle AN. Simultaneous determination of 6 L-arginine metabolites in human and mouse plasma by using hydrophilic-interaction chromatography and electrospray tandem mass spectrometry. *Clin Chem.* 2011; 57: 701-709.
125. Kirsch SH, Herrmann W, Rabagny Y, Obeid R. Quantification of acetylcholine, choline, betaine, and dimethylglycine in human plasma and urine using stable-isotope dilution ultra performance liquid chromatography-tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2010; 878: 3338-3344.
126. Jambor A, Molnar-Perl I. Amino acid analysis by high-performance liquid chromatography after derivatization with 9-fluorenylmethyloxycarbonyl chloride. *J Chromatogr A.* 2009; 1216: 3064-3077.
127. Wang P, Huang L, Davis JL, Swartzman A, Roth B, Stone J, et al. A hydrophilic-interaction chromatography tandem mass spectrometry method for quantitation of serum s-adenosylmethionine in patients infected with human immunodeficiency virus. *Clin Chim Acta.* 2008; 396: 86-88.
128. Yokoyama T, Sokai A, Zenki M. RP-LC with on-line recycled mobile phase for analysis of amino acids pre-derivatized with o-phthalaldehyde. *Chromatographia.* 2008; 67: 535-542.
129. Godejohann M. Hydrophilic interaction chromatography coupled to nuclear magnetic resonance spectroscopy and mass spectroscopy-A new approach for the separation and identification of extremely polar analytes in bodyfluids. *J Chromatogr A.* 2007; 1156: 87-93.
130. Varvara A, Monciu CM, Arama C, Popescu C. Ion-pair reversed-phase high-performance liquid chromatography of ondansetron hydrochlorine using sodium heptanesulphonate as a counterion. *Farmacia.* 2009; 57: 442-451.
131. Kaspar H, Dettmer K, Gronwald W, Oefner PJ. Advances in amino acid analysis. *Anal Bioanal Chem.* 2009; 393: 445-452.
132. De Person M, Chaimbault P, Elfakir C. Analysis of native amino acids by liquid chromatography/electrospray ionization mass spectrometry: comparative study between two sources and interfaces. *J Mass Spectrom.* 2008; 43: 204-215.
133. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens JP, Bouchu D. Ion-pairing reversed-phase liquid chromatography/electrospray ionization mass spectrometric analysis of 76 underivatized amino acids of biological interest: A new tool for the diagnosis of inherited disorders of amino acid metabolism. *Rapid Commun Mass Sp.* 2005; 19: 1587-1602.
134. Bishop MJ, Crow B, Norton D, Paliakov E, George J, Bralley JA. Direct analysis of un-derivatized asymmetric dimethylarginine (ADMA) and L-arginine from plasma using mixed-mode ion-exchange liquid chromatography-tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2007; 859: 164-169.
135. Zotti M, Schiavone S, Tricarico F, Colaianna M, D'Apolito O, Paglia G, et al. Determination of dimethylarginine levels in rats using HILIC-MS/MS: an in vivo microdialysis study. *J Sep Sci.* 2008; 31: 2511-2515.

136. COMMISSION DECISION of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. Official Journal of the European Communities. 2002.
137. Nolin TD, McMenamin ME, Himmelfarb J. Simultaneous determination of total homocysteine, cysteine, cysteinylglycine, and glutathione in human plasma by high-performance liquid chromatography: Application to studies of oxidative stress. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2007; 852: 554-561.
138. Guan X, Hoffman B, Dwivedi C, Matthees DP. A simultaneous liquid chromatography/mass spectrometric assay of glutathione, cysteine, homocysteine and their disulfides in biological samples. *J Pharm Biomed Anal.* 2003; 31: 251-261.
139. Rossi R, Milzani A, Dalle-Donne I, Giustarini D, Lusini L, Colombo R, et al. Blood glutathione disulfide: in vivo factor or in vitro artifact? *Clin Chem.* 2002; 48: 742-753.
140. Refsum H, Helland S, Ueland PM. Radioenzymic determination of homocysteine in plasma and urine. *Clin Chem.* 1985; 31: 624-628.
141. Valerio A, Baldo G, Tessari P. A rapid method to determine plasma homocysteine concentration and enrichment by gas chromatography/mass spectrometry. *Rapid Commun Mass Sp.* 2005; 19: 561-567.
142. Hellmuth C, Koletzko B, Peissner W. Aqueous normal phase chromatography improves quantification and qualification of homocysteine, cysteine and methionine by liquid chromatography-tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2011; 879: 83-89.
143. Mommaerts K, Sanchez I, Betsou F, Mathieson W. Replacing β -mercaptoethanol in RNA extractions. *Anal Biochem.* 2015; 479: 51-53.
144. Cleland WW. Dithiothreitol, a new protective reagent for SH groups. *Biochemistry.* 1964; 3: 480-482.
145. Bylda C, Thiele R, Kobold U, Volmer DA. Recent advances in sample preparation techniques to overcome difficulties encountered during quantitative analysis of small molecules from biofluids using LC-MS/MS. *Analyst.* 2014; 139: 2265-2276.
146. Davey JF, Ersser RS. Amino acid analysis of physiological fluids by high-performance liquid chromatography with phenylisothiocyanate derivatization and comparison with ion-exchange chromatography. *J Chromatogr, Biomed Appl.* 1990; 528: 9-23.
147. Van de Merbel NC. Quantitative determination of endogenous compounds in biological samples using chromatographic techniques. *TrAC-Trend Anal Chem.* 2008; 27: 924-933.
148. Squellerio I, Caruso D, Porro B, Veglia F, Tremoli E, Cavalca V. Direct glutathione quantification in human blood by LC-MS/MS: comparison with HPLC with electrochemical detection. *J Pharm Biomed Anal.* 2012; 71: 111-118.
149. Wang JM, Chu Y, Li W, Wang XY, Guo JH, Yan LL, et al. Simultaneous determination of creatine phosphate, creatine and 12 nucleotides in rat heart by LC-MS/MS. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2014; 958: 96-101.

150. Middtun O, Kvalheim G, Ueland PM. High-throughput, low-volume, multianalyte quantification of plasma metabolites related to one-carbon metabolism using HPLC-MS/MS. *Anal Bioanal Chem.* 2013; 405: 2009-2017.
151. Servillo L, Giovane A, D'Onofrio N, Casale R, Cautela D, Castaldo D, et al. Determination of homoarginine, arginine, NMMA, ADMA and SDMA in biological samples by HPLC-ESI-mass spectrometry. *Int J Mol Sci.* 2013; 14: 20131-20138.
152. Hastie T, Tibshirani R, Friedman J. Linear Methods for Classification. In: Hastie T, Tibshirani R, Friedman J, editors. *The elements of statistical learning.* New York: Springer; 2008; p. 101-118.
153. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen.* 1936; 7: 179-188.
154. Geisser S. Posterior odds for multivariate normal distributions. *J Roy Stat Soc B Met.* 1964; 26: 59-76.
155. Cover TM, Hart PE. Nearest Neighbor Pattern Classification. *IEEE T Inf Theory.* 1967; IT-13: 21-27.
156. Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst.* 2001; 58: 109-130.
157. Szymanska E, Brodrick E, Williams M, Davies AN, van Manen H-J, Buydens LMC. Data Size Reduction Strategy for the Classification of Breath and Air Samples Using Multicapillary Column-Ion Mobility Spectrometry. *Anal Chem.* 2015; 87: 869-875.
158. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics.* 2011; 12: 253.
159. Vanholder R, De SR, Glorieux G, Argiles A, Baurmeister U, Brunet P, et al. Review on uremic toxins: classification, concentration, and interindividual variability. *Kidney Int.* 2003; 63: 1934-1943.
160. Schwartz GJ, Haycock GB, Edelmann CM, Spitzer A. A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine. *Pediatrics.* 1976; 58: 259-263.
161. Schwartz GJ, Munoz A, Schneider MF, Mak RH, Kaskel F, Warady BA, et al. New equations to estimate GFR in children with CKD. *J Am Soc Nephrol.* 2009; 20: 629-637.
162. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Ann Intern Med.* 1999; 130: 461-470.
163. Vervoort G, Willems HL, Wetzels JFM. Assessment of glomerular filtration rate in healthy subjects and normoalbuminuric diabetic patients: validity of a new (MDRD) prediction equation. *Nephrol Dial Transpl.* 2002; 17: 1909-1913.
164. Counahan R, Chantler C, Ghazali S, Kirkwood B, Rose F, Barratt TM. Estimation of glomerular filtration rate from plasma creatinine concentration in children. *Arch Dis Child.* 1976; 51: 875-878.

-
165. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron*. 1976; 16: 31-41.
 166. Way AF, Bolonger AM, Gambertogli JG. Pharmacokinetics and drug dosing in children with decreased renal function. In: Avner ED, Holliday MA, Barratt TM, editors. *Pediatric Nephrology*. 3rd ed. Berlin: Williams&Williams; 1994; p. 1306.
 167. Mihout F, Shweke N, Bige N, Jouanneau C, Dussaule J-C, Ronco P, et al. Asymmetric dimethylarginine (ADMA) induces chronic kidney disease through a mechanism involving collagen and TGF-B1 synthesis. *J Pathol*. 2011; 223: 37-45.
 168. Valli A, Carrero JJ, Qureshi AR, Garibotto G, Barany P, Axelsson J, et al. Elevated serum levels of S-adenosylhomocysteine, but not homocysteine, are associated with cardiovascular disease in stage 5 chronic kidney disease patients. *Clin Chim Acta*. 2008; 395: 106-110.
 169. Jabs K, Koury MJ, Dupont WD, Wagner C. Relationship between plasma S-adenosylhomocysteine concentration and glomerular filtration rate in children. *Metab Clin Exp*. 2006; 55: 252-257.
 170. Romeu M, Nogues R, Marcas L, Sanchez-Martos V, Mulero M, Martinez-Vea A, et al. Evaluation of oxidative stress biomarkers in patients with chronic renal failure: a case control study. *BMC Res Notes*. 2010; 3: 20.
 171. Zinellu A, Sotgia S, Loriga G, Deiana L, Satta AE, Carru C. Oxidative stress improvement is associated with increased levels of taurine in CKD patients undergoing lipid-lowering therapy. *Amino Acids*. 2012; 43: 1499-1507.
 172. Ceballos-Picot I, Witko-Sarsat V, Merad-Boudia M, Nguyen AT, Thevenin M, Jaudon MC, et al. Glutathione antioxidant system as a marker of oxidative stress in chronic renal failure. *Free Radical Biol Med* 1996; 21: 845-853.



CHAPTER V.

Untargeted metabolomics for plasma biomarker discovery using LC-QTOF-MS

- 5.1. Introduction
- 5.2. Untargeted metabolomics workflow
 - 5.2.1. Sampling
 - 5.2.2. Sample preparation
 - 5.2.3. Instrumental analysis
 - 5.2.4. Data processing
 - 5.2.5. Data analysis
 - 5.2.6. Compound identification
- 5.3. Untargeted metabolomics in nephrology: literature review
- 5.4. Study's goal
- 5.5. Materials and equipment
 - 5.5.1. Standards and reagents
 - 5.5.2. Instrumentation
 - 5.5.3. Subjects and sampling
- 5.6. LC-QTOF-MS analytical method
 - 5.6.1. Optimization of protein precipitation
 - 5.6.2. Quality control sample preparation
 - 5.6.3. Instrumental analysis
 - 5.6.4. Data processing
 - 5.6.5. Data analysis
 - 5.6.6. Metabolite identification
- 5.7. Biological interpretation of the results
- 5.8. Conclusions
- 5.9. Bibliography

5.1. INTRODUCTION

CKD is known to be a heterogeneous disease with multiple etiology and complex pathogenesis. This makes difficult to find new sensitive biomarkers that might improve the ability of traditional markers of renal function to detect renal failure regardless of the cause of the disease.

In addition, targeted metabolomics of previously selected metabolites or metabolite groups from a defined metabolic pathway suspicious of being altered was one of the most frequent approaches to find new biomarkers up to the start of the 21st century.

Nowadays, the application of untargeted metabolomics in CKD provides the opportunity to reveal new insights into metabolic profiling and pathophysiological processes¹. Moreover, kidney function has a broad impact on circulating metabolite levels and at the same time, metabolites themselves play functional roles in CKD pathogenesis and in complications. Consequently, untargeted metabolomics approaches have been recognized as particularly promising in nephrology².

5.2. UNTARGETED METABOLOMICS WORKFLOW

Untargeted metabolomics consists of an unbiased holistic approach towards simultaneously analyzing as many features as possible after minimal sample pretreatment, without any prior knowledge of the identity of these features and any bias³⁻⁵. These features include small peptides, amino acids, nucleic acids, carbohydrates, organic acids, vitamins, polyphenols, alkaloids and inorganic species⁶. Analyses are usually carried out semi-qualitatively by determining a subset of the whole metabolite profile⁵. Multivariate statistical analysis becomes indispensable to deal with a much more complex data set in comparison with data from targeted approaches. Moreover, it is essential to find the metabolites that contribute to the differentiation of each particular group of samples. Thereafter, effort is put towards identifying these discriminating molecules⁴.

Untargeted metabolomics are carried out by using either nuclear magnetic resonance (NMR) or mass spectrometry (MS) technologies coupled to chromatographic separation. The combination of mass spectrometry with a separation technique has several advantages, because it reduces the complexity of mass spectra and gives additional

information regarding the physicochemical properties of the analytes, and at the same time, offers high selectivity, sensitivity and improves the potential to identify metabolites⁷.

Regardless the technique of choice, studies following untargeted metabolomics require several well-differentiated steps: sampling, storage of samples, sample pretreatment, system conditioning, sample injection, data pre-processing of the resulting features, data analysis, compound identification and results interpretation⁸. A careful forward planning of all of these steps becomes essential for untargeted metabolomics. Figure 5.1 shows an example of a LC-MS based untargeted metabolomics workflow.

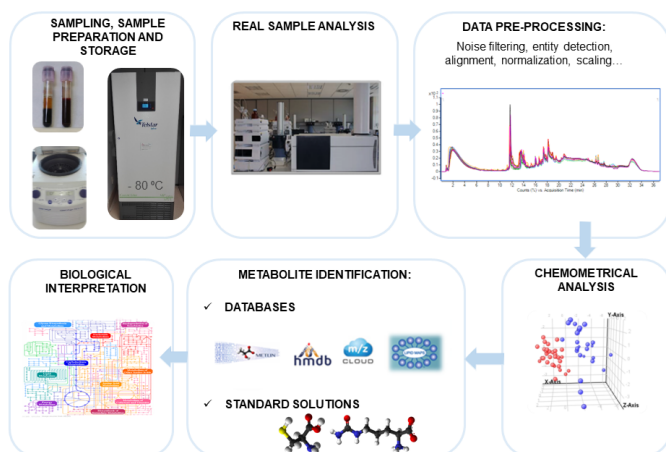


Fig. 5.1. Typical LC-MS untargeted metabolomics workflow.

Complex multivariate data sets are obtained from untargeted metabolomics studies, thus chemometric tools play a key role for data analysis and interpretation. In addition, omics databases contain up-to-date information about physical and chemical properties of a high amount of molecules, data about their function, localization and expression, information regarding similar experiments already carried out and experimental data, such as spectra. For that reason, omics databases are invaluable instruments for the identification of significant entities obtained from chemometric approaches in untargeted metabolomics analyses and have already been remarked as inestimable tools in nephrology research⁹.

5.2.1. SAMPLING

Operations involved in both sampling and sample preparation procedures might affect and even modify the composition of the measured metabolome, leading to differences between displayed metabolome and real metabolome composition at the time of sampling¹⁰. Therefore, sampling procedures must be carefully studied and carried out.

Traditionally, blood sampling has prevailed rather than collection of urine or other body fluids for metabolomics studies due to its clinical importance and practical reasons. Indeed, the chemical composition of blood reflects tissue lesions, organ dysfunctions and pathological states, and the collection and storage of blood samples is usual both in biobanks and for clinical test purposes^{11,12}. However, it is also possible to use other body fluids including urine, saliva, cerebrospinal fluid, feces or follicular fluid¹¹.

Regarding blood sampling, it has to be noted that there are differences in composition of plasma and serum. The protein content in serum is lower because the proteins involved in coagulation are removed. However, the blood clotting process performed to obtain serum generally takes place at room temperature during 30-60 minutes, which might increase metabolite degradation processes, might increment the enzyme conversion rates and some metabolites could be lost during clot precipitation¹⁰. One of the advantages of plasma is that during plasma formation the problem of keeping blood samples at room temperature is avoided, due to the fact that this process is carried out using a centrifuge at 4°C. Nevertheless, blood to plasma transformation requires the use of anticoagulant containing sample tubes. This means that the effect on the metabolomics profile of the anticoagulants used in the tubes for blood collection to obtain plasma (such as, EDTA, lithium heparin, sodium citrate or sodium fluoride) needs to be studied¹³.

Recent research shows that the metabolite profiles from plasma and serum might not be so different. The main differences found for some features are related to retention time shift (which makes difficult to assure whether they correspond to the same metabolite or not), differences in multiple-charge ions corresponding to peptides and differences in the adducts formed during ionization^{14,15}. Metabolite concentration in plasma and serum usually keeps a good correlation, even though direct extrapolation of the results of plasma to serum and vice versa is not advisable^{16,17}.

The election of plasma or serum is a pragmatic choice that depends on personal preferences and on the available equipment. In any case, it is of great importance the use

of only one sample type for each particular untargeted metabolomics study¹⁷. Despite the fact that some noteworthy studies, such as Human Metabolome Project and HUSERMET have been carried out in serum, plasma accounts for the most prevalent biofluid used in LC-MS untargeted metabolomics. Indeed, plasma is used in the 65 % of studies, whereas 20 % of the researches are done with serum^{10,18}.

5.2.2. SAMPLE PREPARATION

It is estimated that the most frequently used biofluids, blood and urine, may contain almost 5000 metabolites across different chemical families in a large range of concentrations. Therefore, sample preparation methods should be simple, nonselective, fast and reproducible^{12,19}. The typical main steps for sample preparation in metabolomics include: quenching and deproteinization or "dilute-and-shoot" step (depending upon the biofluid)¹⁹.

Blood collection in the context of omic studies needs for a safety margin definition for sample handling after blood drawing as well as the establishment of a procedure to ensure sample quality¹¹. Indeed, there is increased interest in the shortening of the time window between sampling and analysis, as the turnover kinetics of some metabolites is known to be particularly fast²⁰. For that reason, with the aim of avoiding the effect of enzymatic action, rapid inactivation of metabolism or quenching is the first step immediately after sampling²¹.

Quenching step becomes indispensable in metabolomics to ensure that the sample represents the true composition of the metabolome at the time of sampling¹⁰. The most frequent strategies consist in the rapid modification of sample conditions, usually pH or temperature²². The modification of pH is achieved through the addition of basic solutions (e.g. KOH or NaOH) or extremely acid ones (e.g. perchloric acid, HCl or trichloroacetic acid)^{13, 15}. However, the use of extreme pH is not frequent as it may result in a severe reduction of the number of metabolites detected as well as in degradation of unstable metabolites affected by pH²². The modification of temperature is more common for quenching purposes, because it is generally assumed that the shock caused by sample cooling under -20 °C does not affect the integrity of the sample^{13, 15}. Moreover, if samples are immediately frozen under -80°C, the majority of the metabolites are preserved²⁰. The next step after quenching depends on the biofluid used for untargeted metabolomics. When working with urine it is common to dilute the samples with an aqueous solution using

a dilution factor between 1:1 and 1:10 to decrease potential problems of ion suppression (known as “dilute-and-shoot” step)²³. On the other hand, blood contains around 6-8 g/dL of proteins, which makes necessary protein removal when working with serum or plasma and also with tissues²⁴. Protein removal is usually carried out by means of protein precipitation methods (PPT), adding different solvents in plasma:precipitant agent ratios between 1:2 and 1:4²⁵⁻²⁷ (see Fig. 5.2). Acetonitrile, methanol, ethanol and acetone alone or in combination are some of the most frequently used solvents for this purpose^{25,28,29}. It has to be considered that the addition of organic solvent to biofluids also works as disruptor of the linkages between metabolites and proteins present in the solution¹⁰. Moreover, these solvents operate not only as protein precipitation agents, but also as metabolite extractors²⁷.

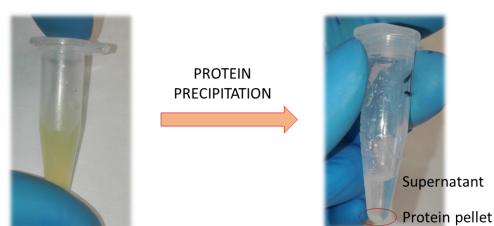


Fig. 5.2. Plasma sample before and after protein precipitation by means of the addition of precipitation agent and centrifugation.

5.2.3. INSTRUMENTAL ANALYSIS

A wide range of analytical instruments might be applied in untargeted metabolomics research. Nuclear magnetic resonance (NMR) or mass spectrometry (MS) technologies coupled to chromatographic separation techniques, such as liquid chromatography (LC) and gas chromatography (GC), are the most usual analytical techniques for untargeted metabolomics.

There is not any universal analytical technique for untargeted metabolomics; all of them have their advantages and drawbacks, as previously explained in Chapter 3. Due to the diversity and broad dynamic range of metabolites in body fluids, for a more global approach, the use of multi-analysis techniques is recommended to overcome the shortcomings of different single analysis techniques^{5,6}.

Even though NMR and GC were pioneering techniques for metabolomics, the use of LC-MS for global metabolite profiling is increasing³⁰. Regarding the use of different mass

analyzers in LC-MS untargeted metabolomics, time-of-flight (TOF), Orbitrap and Fourier-transform ion cyclotron resonance (FT-ICR) are the most popular instruments.

5.2.4. DATA PROCESSING

Untargeted metabolomics studies generate complex multivariate data sets, which need for data processing steps prior to data analysis. These data sets generally contain complex three-dimensional information: retention time, mass-to-charge ratio and intensities of the features. For that reason, data processing workflow is complex and needs for different software to comply with the following steps:

1. Raw signal filtering to remove noise and baseline correction
2. Feature detection
3. Chromatographic and spectral data alignment to correct differences in retention times and exact mass between runs
4. Normalization steps for systematic variation between samples in terms of abundance prior to data analysis.

These steps involve several parameters that can be optimized, being the quality of the processed data and the number of peaks obtained dependant on these settings.

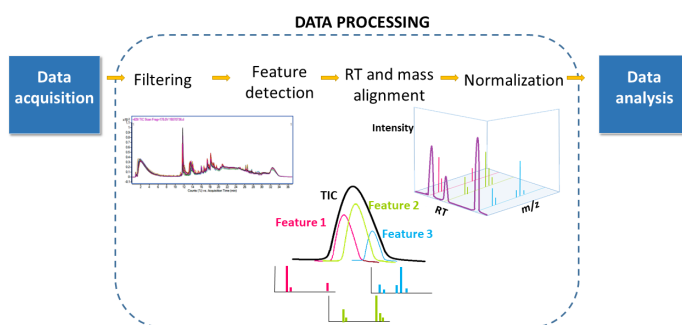


Fig. 5.3. Data processing workflow in untargeted metabolomics.

Data processing might be carried out by means of several commercial and open source software containing routines for automatic alignment, denoising, deconvolution and extraction of peaks³¹. Different possibilities exist for data processing:

- Using vendor's commercial software (such as, MarkerLynx, Markerview, Profiler, MassHunter Qualitative, Genespring, etc.)
- Using software from independent developers able to work with all type of MS data files (e.g. GeneData)
- Working with open access programs (for instance, XCMS, MZmine, Metalign, etc.)
- Developing an own script in computing language (typically in Matlab or R)

The used peak picking software itself can have a significant effect on the result of a study. The most practiced option by laboratories is the use of vendor's commercial software. In fact, these software are usually integrated with the MS instrument operating system and often incorporate documentation or tutorials with a wealth of detail and basic to more advanced utilities for data mining³².

5.2.5. DATA ANALYSIS

Data analysis makes use of statistics and chemometrics to get a reasonable number of relevant features from the obtained complex data set that might improve the differentiation between control and CKD samples.

Different chemometric tools are likely to be applied to untargeted metabolomics studies such as the following:

- Principal Component Analysis (PCA)
- Partial Least Square Discrimination Analysis (PLS-DA)
- Sparse Partial Least Square Discrimination Analysis (SPLS-DA)

The first informative look at the data structure and its relationship between groups is given by application of PCA. The ideal would be that the results from PCA analysis would lead to an initial biological conclusion, which would be afterwards tested by means of PLS-DA. Indeed, being PLS-DA a supervised multivariate method, it is more prone to produce biologically relevant results. PLS-DA is used to get the resulting performance from the classification of samples into control or disease groups according to the models built with the selected features with prior knowledge of sample classification. In addition, SPLS-DA is an advantageous and relatively new tool especially useful in untargeted metabolomics for variable selection, and thus for data mining^{33,34}.

5.2.6. COMPOUND IDENTIFICATION

Metabolite identification in untargeted metabolomics is based in the first instance on the comparison of experimental data of the significant entities with lab-built or web-based spectral databases. However, it has to be noted that identification in LC-MS based metabolomics depicts more complexity in comparison to GC-MS. Indeed, the available information in LC-MS libraries is limited to exact mass values, whereas in GC-MS libraries spectra comparison is performed. LC-MS databases usually generate lists of known compounds with their exact mass for the experimental accurate mass values obtained. Some libraries also contain additional information or tools for further biological interpretation. Some of the databases used for compound identification and biological meaning in untargeted metabolomics are listed below³⁵:

- ✓ General chemistry databases including natural and synthetic chemical products, like Scifinder Scholar (<https://scifinder.cas.org/scifinder>) or ChemSpider (<http://www.chemspider.com>)
- ✓ Human metabolome databases, such as, Human Metabolome database (HMDB, www.hmdb.ca) or Metlin (<https://metlin.scripps.edu>)
- ✓ Databases for specific subsections of the metabolome like Lipid Maps Structure Database (<http://www.lipidmaps.org/resources/resources.html>).
- ✓ Mass spectral databases like NIST (<https://www.nist.gov>), mzCloud Advanced Spectral Database (<https://www.mzcloud.org>) or MassBank (<http://www.massbank.jp>)
- ✓ Databases on metabolic pathways, e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg>)

Once the list of potential candidates for each entity is generated, MS/MS analysis provides extra information that may help elucidating the identity of the feature by comparison of reference MS/MS spectra of the metabolites and experimental spectra obtained from the analysis of samples. Finally, unequivocal identification is achieved by means of the comparison of retention time, MS and MS/MS spectra of commercial standards with the experimental data collected.

Chemical Analysis Working Group (CAWG) from the Metabolomics Standards Initiative (MSI) has already established compound identification confidence by classifying the identified compounds in 4 different levels³⁶:

1. Identified compounds (with reference analytical standards)
2. Putatively annotated compounds (based on spectral or physicochemical similarity but without chemical reference standards)
3. Putatively characterized compound classes (based on spectral or physicochemical similarity to known chemical classes)
4. Unknown compounds

5.3. UNTARGETED METABOLOMICS IN NEPHROLOGY: LITERATURE REVIEW

Different studies have been performed to find new biomarkers in nephrology field using an untargeted metabolomics approach, as showed in Table 5.1.

Various untargeted metabolomics experiments have been carried out in adults with CKD³⁷⁻⁴². Regarding paediatric population, an experimental model using newborn rats in acute kidney injury (AKI)⁴³ and two more studies in human paediatrics suffering from nephrouropathies⁴⁴ and AKI⁴⁵ have been found in literature. These untargeted metabolomics studies performed in paediatrics used urine matrix, even though in the experiments carried out in adults both plasma^{38,39,42,46,47} and urine^{40,41,48} matrices are used in a similar proportion. In addition to these matrices, serum³⁷ and hemodialysate sample⁴⁹ have been utilized in untargeted metabolomic approaches as well.

Concerning the analytical technique used, the most commonly used equipment include: ¹H-NMR ^{39,41,44,48,50}, LC-MS ^{38,43,45-47,49} and GC-MS equipment^{37,38,40,43}. Besides, CE-TOF-MS⁴² was also used in another experiment.

To sum up, as far as we know, only a few of the untargeted metabolomics approaches were carried out in paediatrics, and none of them were focused on paediatrics with CKD, where there is a research gap.

Table 5.1. Literature review of untargeted metabolomics methods related to renal area.

Population	Disease	Matrix	Analytical method	Author	Year
Mice	Diabetic nephropathy	Urine	¹ H-NMR	SteC ⁵⁰	2015
Adults	Advanced-stage CKD	Urine	¹ H-NMR	Posada-Ayala ⁴⁸	2014
African American adults	CKD	Serum	GC-MS	Yu ³⁷	2014
Newborn rats	Aminoglycoside induced acute kidney injury (AKI)	Urine	LC-LCO-MS GC-MS	Hanna ⁴³	2013
Adults	CKD	Plasma	UPLC-LTQ-MS GC-MS	Shah ³⁸	2013
Adults	CKD stages 3-4	Plasma	¹ H-NMR	Mutsaers ³⁹	2013
Adults	Diabetes mellitus and CKD	Urine	GC-MS	Sharma ⁴⁰	2013
Young adults	IgA nephropathy	Plasma	LC-QTRAP-MS	Zivkovic ⁴⁶	2012
Adults born with Extremely Low Birth Weight (ELBW)	CKD	Urine	¹ H-NMR	Atzori ⁴¹	2011
Adults on hemodialysis	Chronic glomerulonephritis and diabetic nephropathy	Plasma	LC-QTOF-MS	Sato ⁴⁷	2011
Adults	CKD	Plasma	CE-TOF-MS	Toyohara ⁴²	2010
Paediatrics	Nephrouropathies	Urine	¹ H-NMR	Atzori ⁴⁴	2010
Adults on hemodialysis	Uremic patients	Hemodialysate sample	LC-LTQ/Oorbitrap-MS	Godfrey ⁴⁹	2009
Paediatrics	AKI as a complication of cardiopulmonary surgery	Urine	LC-TOF-MS	Beger ⁴⁵	2008

5.4. STUDY'S GOAL

The main goal of the present work deals with the interest of finding potential biomarkers that could help differentiating between control and CKD paediatric patients without any prior information of which metabolic pathways may be affected and any clue of the metabolites that may be up- or down-regulated.

To accomplish this purpose, an untargeted metabolomics study needs to be carried out using LC-QTOF-MS analytical technique in order to find significant differences between two populations: paediatrics suffering from CKD and control paediatrics.

Both extraction method and analysis method optimization are required to obtain as many entities as possible introducing the least possible uncertainty to the measure. Then, the use of different commercially available software for automatic data processing, as well as for multivariate statistical analysis are needed. Finally, the identification of significant entities using different available tools is required.

5.5. MATERIALS AND EQUIPMENT

5.5.1. STANDARDS AND REAGENTS

Methanol, ethanol and acetonitrile used in the plasma extraction study were acquired from Scharlab (Barcelona, Spain) and dry acetone was purchased from Panreac (Barcelona, Spain). In addition, LC-MS grade formic acid used for mobile phase was obtained from Fisher Scientific (Ghent, Belgium). Ultra-high purity water achieved from tap water pretreated by Elix reverse osmosis and a Milli-Q system from Millipore (Bedford, MA, USA) was used. Moreover, mobile phases were filtered through 0.1 µm filters from Millipore Omnipore (Watford, Ireland) prior to use.

Becton Dickinson EDTA tubes (Plymouth, UK) were used to collect blood samples from both paediatric CKD patients and controls.

After prior identification of the significant features, the following analytical standards were used for identity confirmation purposes: D-erythro-sphingosine-1-phosphate from Larodan AB (Limhamn, Sweden), *cis*-4-decenoylcarnitine synthesized by Lumila Research group at Universidad Autónoma de Madrid (Madrid, Spain), bilirubin from TCI (Tokyo, Japan), and

DL-2-aminoadipic acid, *iso*-butyryl-L-carnitine and *n*-butyryl-L-carnitine from Sigma-Aldrich (Steinheim, Germany).

5.5.2. INSTRUMENTATION

Similar instrumentation to that described in Chapter 4 was used for this untargeted metabolomics study. The analysis was accomplished injecting 5 µL of sample in a 1200 series HPLC system coupled to a 6530 Series hybrid quadrupole time-of-flight mass spectrometer (QTOF) from Agilent Technologies (Santa Clara, CA, USA), equipped with Agilent Jet Stream electrospray source. Chromatographic separation was carried out using the reversed phase column Zorbax SB-C18 (2.1x100 mm, 3.5 µm) and C8 guard column (2.1x12 mm, 5 µm), both from Agilent Technologies, which were kept at 35 °C during the analysis.

Data acquisition was carried out using Agilent MassHunter Workstation Data Acquisition version B.05.01 software and afterwards, raw data was processed using MassHunter Qualitative Analysis version B.07.00 software and Profinder version B.06.00 software (from Agilent Technologies). Finally, data analysis was performed using Mass Profiler Professional software (MPP) version B.13.00 (Agilent Technologies, Santa Clara, CA, USA) and Matlab R2015a (Mathworks, Natick, Massachusetts, USA).

Samples were centrifuged with a 5415 R Eppendorf (Madrid, Spain) centrifuge, and supernatant evaporation was carried out in a Techne, Dri-Block® DB-3D (Staffordshire, UK) evaporator system.

5.5.3. SUBJECTS AND SAMPLING

Thirty-two patients suffering from CKD, aged 3-18 years, and twenty-six control patients, aged 6-19 years were recruited for the study. Paediatrics suffering from CKD were at different degrees of the disease and some of them received renal replacement therapy (RRT). Characteristics of the patients are showed in Table 5.2.

Table 5.2. Characteristics of the paediatric population subjected to study.

Characteristics of the population				Number of patients
CKD DEGREE	SEX (M/F)	AGE (2-12 y/13-18 y)	TREATMENT (No RRT/ Dialyzed/ Transplanted)	
CONTROL	20/6	15/11	26/0/0	26
CKD2	8/6	10/4	9/0/5	14
CKD3	5/1	2/4	4/0/2	6
CKD4	2/4	3/3	5/0/1	6
CKD5	2/4	5/1	1/5/0	6

All paediatric CKD patients were followed up in Cruces University Hospital (Bilbao, Spain) and were clinically stable at the time of study. CKD samples were the same used in Chapter 4, whereas some control samples differed from one study to another. Patients with anuria, hepatopathy and/or insulin-dependent diabetes mellitus were discarded, as previously described in Chapter 4. In addition, control patients met the following inclusion criteria: healthy children who had a minor surgery in Cruces University Hospital.

Blood samples were collected in EDTA tubes in the morning after an overnight fasting. Immediately after blood collection, samples were cooled in an ice-water bath and centrifuged at 1000 g for 5 min at 4°C. Finally, samples were stored at -80°C until sample treatment and analysis.

Ethics Committee of Clinic Research of Cruces Hospital approved the study protocol and patients' parents gave written informed consent.

5.6. LC-QTOF-MS ANALYTICAL METHOD

5.6.1. OPTIMIZATION OF PROTEIN PRECIPITATION

Minimal sample treatment is recommended in untargeted metabolomics to perform the extraction of as many metabolites as possible without any bias. The use of different extractant combinations for plasma protein precipitation and metabolite extraction has been investigated in order to get features with different polarities: acetonitrile, methanol, ethanol, acetonitrile:methanol (1:1, v/v), methanol:ethanol (1:1, v/v) and acetone:methanol (1:1, v/v). Plasma and precipitation solvent ratio of 1:3 (v/v) was selected according to our previous experience and literature available about plasma protein precipitation^{29,51-54}. Plasma precipitation consisted in taking 50 µL plasma and adding 150 µL of the selected extractant (n=5). Then, samples were vortexed, stored at

-40 °C for 5 min and subsequently centrifuged at 13000 rpm for 10 min at 4 °C. Finally, the supernatant was injected in the LC-QTOF system.

These experiments were carried out with the following preliminary chromatographic conditions. A binary solvent system consisting of 0.1 % formic acid with 3 % acetonitrile (A) and acetonitrile with 0.1 % formic acid and 5 % water (B), at a flow rate of 300 µL/min, with the following gradient program: 20 % B was kept for 2 minutes, a linear increase of organic phase was performed to reach 100 % B in 22 minutes, 100 % B was kept for 2 minutes, gradient returned to 20 % B in 2 minutes and finally, 20 % B was kept for 5 minutes.

First of all, chromatographic profile obtained from the analysis of the selected solvents was assessed. Acetone:methanol solvent combination was discarded for protein precipitation, due to the fact that the TIC of blank acetone:methanol analysis contained a high number of ups and downs or small peaks (see Fig. 5.4) and this might introduce considerable uncertainty in the analysis.

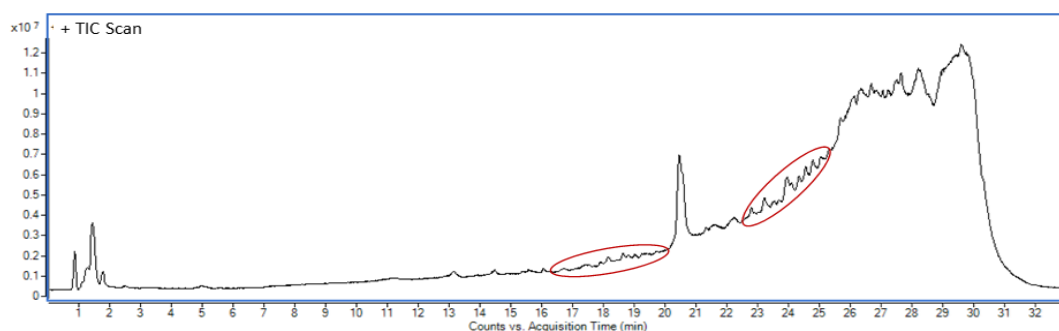


Fig. 5.4. Example of acetone:methanol blank sample TIC chromatogram. In red, details of the uncertainty introduced.

After discarding the use of acetone:methanol as precipitation solvent, acetonitrile, methanol, ethanol, acetonitrile:methanol (1:1, v/v) and methanol:ethanol (1:1, v/v) were considered as precipitation reagents and plasma protein precipitation and extraction of the metabolites was carried out as stated before in pooled plasma.

Data processing for feature extraction of different plasma extracts was performed using preliminary not very restrictive conditions for MFE using MassHunter Qualitative software (minimum peak abundance ≥ 1800 counts, absolute height ≥ 5000 counts, quality score ≥ 70 and the following charge states: H^+ , Na^+ , K^+ and neutral loss of H_2O). Then, PCA of the features achieved for the different aliquots of pooled plasma was obtained and pooled plasma samples extracted with different solvent combinations were coloured accordingly.

The results reveal that there is a clear separation between extracts using different solvents and some degree of separation is found between solvent mixtures, as showed in Fig. 5.5. This means that the extracted entities depend to a great extent on the solvent or solvent mixture used for protein precipitation.

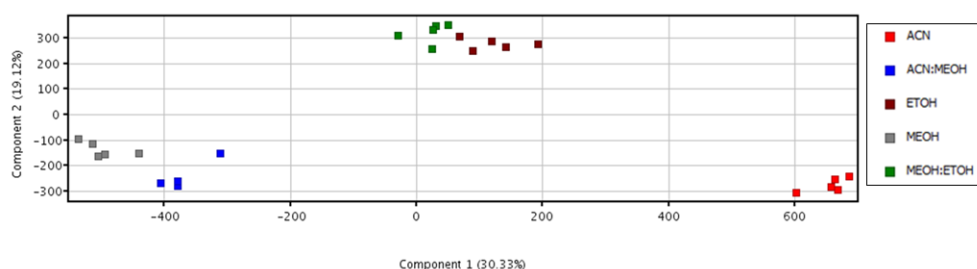


Fig. 5.5. PCA of plasma sample extracts depending on the solvent used for protein precipitation: acetonitrile (ACN), acetonitrile:methanol (ACN:MEOH), ethanol (ETOH), methanol (MEOH) and methanol:ethanol (MEOH:ETOH).

Once checked that the extraction solvent lead to different results in feature extraction, some other concerns were taken into account to select the optimum solvent for protein precipitation. With the aim of defining the ideal extractant to maximize feature extraction from plasma, the number of entities obtained for each solvent was assessed. Table 5.3 shows that the amount of extracted entities were in the same range, except for acetonitrile, which had a lower number of features extracted.

Table 5.3. Mean number of features obtained from plasma extraction using different solvents and standard deviation (n=5).

Solvent	Mean features ± standard deviation
Acetonitrile	686 ± 16
Acetonitrile:methanol (1:1, v/v)	769 ± 33
Methanol:ethanol (1:1, v/v)	773 ± 12
Methanol	778 ± 11
Ethanol	802 ± 28

In addition, visual inspection showed that pellets obtained from protein precipitation with ethanol were not as consistent as the pellets achieved with the rest of protein precipitations. Therefore, ethanol was ruled out because this could be in relation with a lower rate of protein removal and could be a cause of protein accumulation in the chromatographic column reducing its useful life.

The effect of injecting consecutive extracts obtained from protein precipitation of pooled plasma with these solvents was evaluated as well. Regarding the use of acetonitrile for

protein precipitation, the analysis of subsequent plasma quality control (QC) samples extracted with this solvent reveals an increase in TIC chromatogram (see Fig. 5.6).

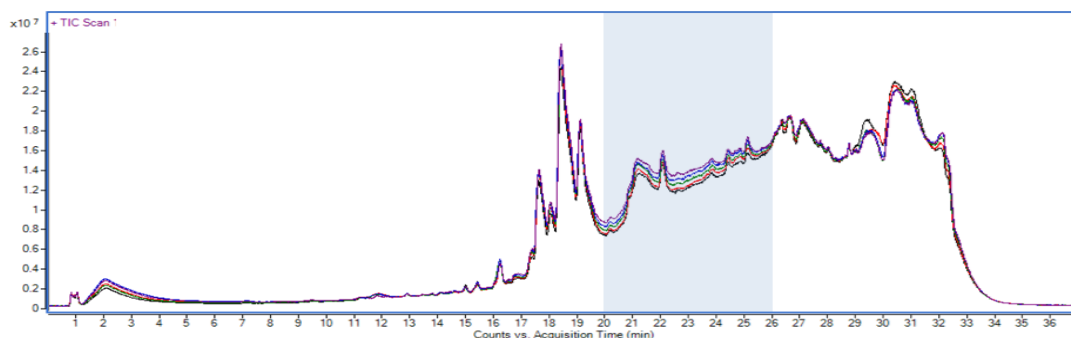


Fig. 5.6. Subsequent TIC chromatograms of a QC plasma extract using acetonitrile as protein precipitation solvent. Blue shading shows the increase of TIC.

The study of the increased ions from consecutive acetonitrile extract analysis showed that there was a phospholipid accumulation, as identified by comparing the experimental spectra obtained with spectra from Postle et al⁵⁵. The correspondence of the m/z values with their respective lipids is showed in Figure 5.7. A higher extraction of phospholipids from plasma was suggested for acetonitrile in comparison with other extractants, which tend to accumulate in the chromatographic column. Consequently, the use of acetonitrile as extractant solvent, alone or in a mixture, was discarded.

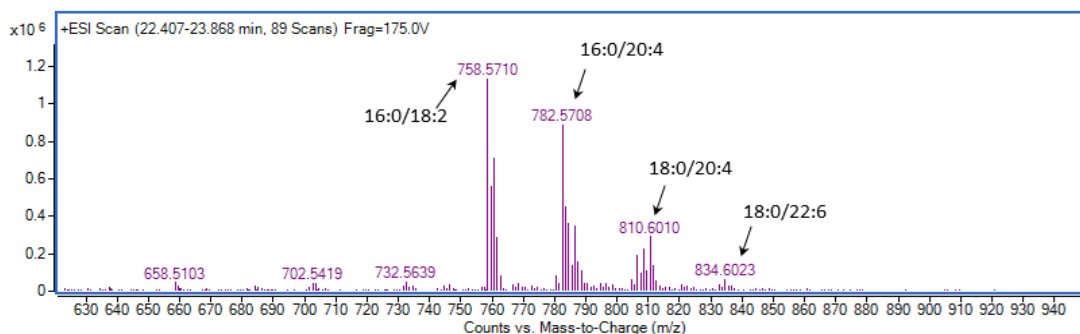


Fig. 5.7. Subsequent TIC chromatograms of a QC plasma extract using acetonitrile as protein precipitation solvent.

After discarding acetonitrile, acetonitrile:methanol (1:1, v/v), ethanol and acetone:methanol (1:1, v/v), being the number of entities alike for methanol and methanol:ethanol, the shape of the chromatographic peaks was also considered to choose one of them. Finally, due to the fact that no significant difference was found between different extracts, in accordance with previous experience in untargeted

metabolomics^{51,53,56} and being methanol:ethanol an intermediate mixture, methanol:ethanol (1:1, v/v) was selected.

Therefore, sample preparation consisted in the following procedure. Plasma was thawed at room temperature and 50 μ L of each sample were placed in Eppendorf tubes. In addition, 150 μ L of frozen methanol:ethanol (50:50) were added for protein precipitation. Then, samples were vortexed, stored at -40 $^{\circ}$ C for 5 min and centrifuged at 13000 rpm for 10 min at 4 $^{\circ}$ C, as showed in Fig. 5.8.



Fig. 5.8. Sample treatment for plasma protein precipitation and metabolite extraction.

5.6.2. QUALITY CONTROL SAMPLE PREPARATION

Pooled plasma was prepared to be used as quality control samples (QC). 50 μ L of each control sample were pooled in a Falcon tube. The same operation was carried out with all the CKD samples in a different Falcon tube. Finally, equal quantities of both pooled plasma were mixed, as showed in Figure 5.9.

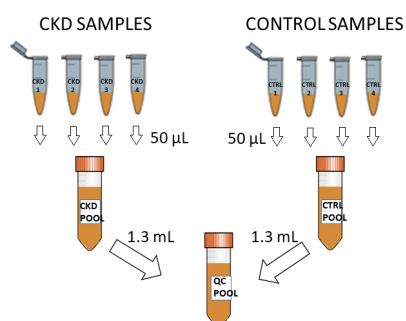


Fig. 5.9. QC preparation pooling plasma from CKD and control samples.

QC samples received the same sample treatment as the samples (see Fig. 5.8) and were injected at the beginning of the run and randomly between samples during sequence analysis.

5.6.3. INSTRUMENTAL ANALYSIS

Chromatographic separation was performed under gradient mode. It is desirable that the chromatogram covers the maximum polarity range, by starting in a low percentage of organic phase and ending with a high percentage. Gradients starting from 5 %, 10 %, 15 % and 20 % of organic phase (B) were assayed. Total ion current (TIC) chromatograms of gradients starting from 5 % to 15 % B presented uncertainty from minutes 7 to 10 of the chromatogram, which was reduced as the percentage of organic phase increased. However, when reaching 20 % B at the start of the gradient, the uncertainty disappeared. Thus, a gradient covering 20 % B to 100 % B was preferred for the untargeted metabolomics study.

Once selected the initial proportion of the binary solvent system, the gradient was adapted to plasma extract analysis, by elongating the time in which 100 % of organic phase was supplied from 2 to 4 minutes, in order to avoid the accumulation of apolar compounds (see Table 5.4).

Table 5.4. Gradient analysis used previously for blank solvent samples (left) and adaptation of the gradient for plasma extracts (right).

Prior gradient for blank solvents		Adaptation of gradient to plasma extracts	
Time (min)	%B	Time (min)	%B
0	20	0	20
2	20	2	20
24	100	24	100
26	100	28	100
28	20	30	20
33	20	37	20

Between different analytical runs sample wash was performed rinsing the needle for 30 s using a flushport with acetonitrile:isopropanol (60:40, v/v) and 0.2 % formic acid.

The LC system was coupled to the mass spectrometer with an ESI ionization source in positive mode. The mass spectrometer worked in MS positive mode and mass spectra were acquired scanning over 100-1000 m/z range with an acquisition rate of 1 spectra/s. Preliminary studies showed that it was feasible not cleaning the ion source during the analysis as system conditions kept constant. Indeed, the number of samples was not too large to dirty the interface and disturb the ionization.

Both the definitive conditions for chromatographic separation and mass spectrometry analysis are summarized in Table 5.5.

Table 5.5. LC-QTOF-MS conditions for the metabolomic study.

Chromatographic conditions		Mass spectrometry conditions	
Column	Zorbax SB-C18 (2.1x100 mm, 3.5 μ m) C8 guard column (2.1x12.5 mm, 5 μ m)	Interface	ESI Agilent Jet Stream®
Column temperature	35 \pm 0.8 $^{\circ}$ C	Nozzle	0 V
Mobile phase	A: 0.1 % formic acid with 3 % acetonitrile B: acetonitrile with 0.1 % formic acid and 5 % water	Fragmentor voltage	175 V
Flow	0.3 mL/min	Dry gas	300 $^{\circ}$ C, 10 L/min
Gradient program	0 min, 20 %B	Nebulizer pressure	20 psi
	2 min, 20 %B	Sheath gas	350 $^{\circ}$ C, 11 L/min
	24 min, 100 %B	Vcap	4,000 V
	28 min, 100 %B	Skimmer	65 V
	30 min, 20 %B	OCT 1 RF	750 V
37 min, 20 %B			
Injection volume	5 μ L, with needle wash	Mass range	MS: 100-1000 m/z MS/MS: 30-1000 m/z

An example of TIC chromatogram of a QC and a blank sample (methanol:ethanol 1:1, v/v) acquired under these conditions is showed in Fig. 5.10.

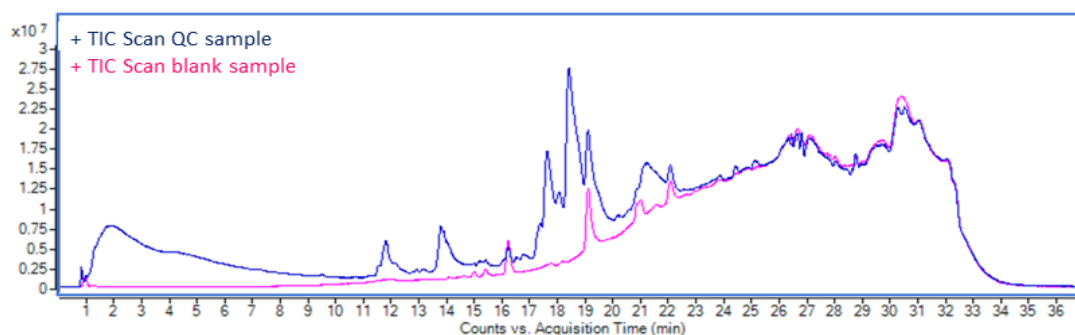


Fig. 5.10. Example TIC chromatograms of QC and blank (methanol:ethanol 1:1, v/v) samples.

A relevant aspect to be considered in instrumental analysis of untargeted metabolomics studies is that there is a need for assuring the mass accuracy of the recorded ions and mass drift compensation by continuous internal mass calibration during the analyses. For that purpose, reference ions 121.0509 m/z (purine) and 922.0098 m/z (HP-921) were monitored. In addition, during the analysis, every 12 hours external calibration was also performed and before following with plasma sample analysis, two no injection and two QC samples were analyzed.

In addition, all samples were randomly injected along the sequence and immediately after the analysis, samples were kept frozen to carry out MS/MS analysis after the chemometric analysis.

5.6.4. DATA PROCESSING

The first procedure after analyzing plasma extracts by LC-QTOF-MS consisted in checking the obtained raw chromatographic and spectral data using MassHunter Qualitative Analysis version B.07.00 software. After this initial step, data was processed using Profinder version B.06.00 software (both from Agilent Technologies).

First of all, pressure variability, baseline drift, reference mass infusion and retention time shift were monitored and verified by means of MassHunter Qualitative Analysis software, in order to control the inherent variation usually related to chromatographic process.

It has to be noted that between-sample pressure variability could mean that some uncertainty has been produced during the analysis. For that reason, pressure superposition of different plasma samples was performed to check that the chromatographic system did not suffer from alterations. Figure 5.11 shows that no variation of pressure was observed for plasma sample analysis.

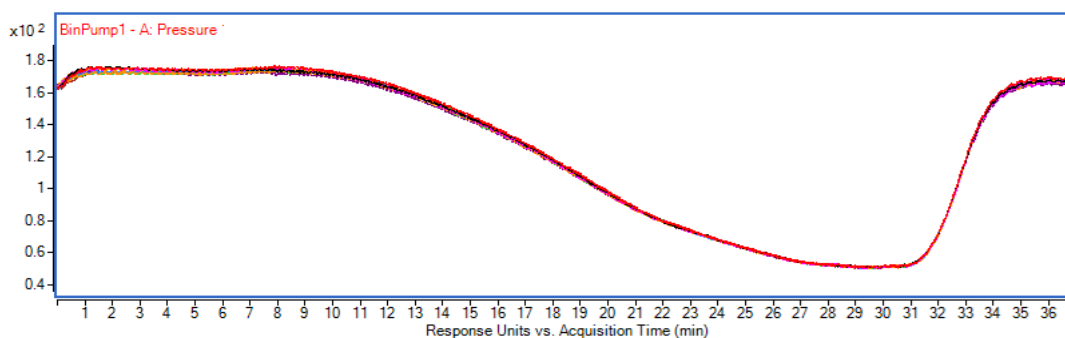


Fig. 5.11. Pressure superposition of different plasma samples.

In addition, the accumulation of compounds present in plasma extract would lead to TIC baseline increase. Similarly, the dirty status of the interface could contribute to signal suppression thus causing baseline decrease. The stability of the baseline was checked by superimposing different CKD and control plasma samples. Plasma samples showed a stable baseline, despite containing, as expected, different peaks along the TIC (see Fig. 5.12).

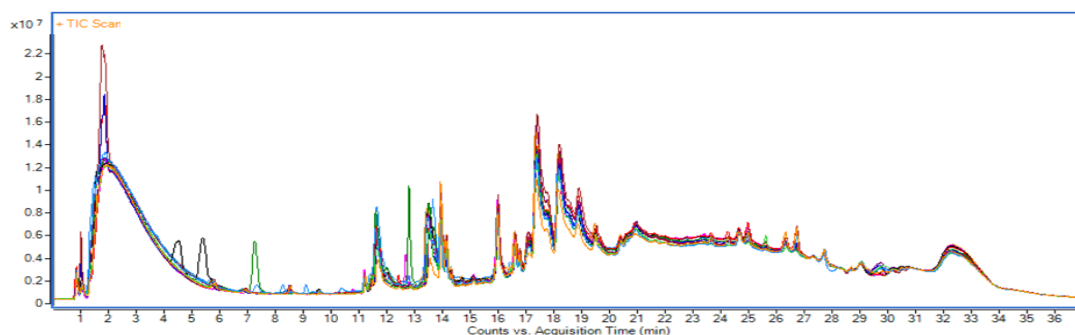


Fig. 5.12. TIC chromatogram superposition of different samples, with matching baselines.

Moreover, to assure that reference mass correction by means of reference standard infusion had performed correctly, the sufficient abundances of 121.0509 m/z and 922.0098 m/z ions used for mass correction were checked by monitoring extracted ion chromatograms (EIC). EIC chromatograms extracted with an error of 30 ppm showed constant signal for both reference standard ions.

Once checked the apparent quality of the data acquired, Recursive Feature Extraction (RFE) algorithm was performed using Profinder software. This algorithm includes Molecular Feature Extraction (MFE), which intends to remove chemical background and rapidly find feature peaks in TIC chromatogram taking into account isotope distribution. Then, MFE is followed by retention time and mass alignment across sample data and integration of the corresponding chromatographic peaks. In addition, RFE algorithm creates a small list of unique features after time and mass alignment in addition to further filters used.

The filters selected in Profinder define to a large extent the number of features extracted. First of all, less restrictive filters were used and once studied the background, peak shape, number of ion species defining each peak and abundance, the definitive filters were selected. Ultimate filters were set as follows: m/z range (100-1000 m/z), peak height (>1500 counts), absolute peak height (>10000 counts), ion species (protonated ion, sodium adduct, potassium adduct and neutral loss of water), charge state (set to a maximum of 2, thus allowing the inclusion of small molecules and lipids), maximum exact mass (>1000 Da), peak spacing tolerance (0.0025 m/z , plus 7 ppm), MFE score (70 %), RT alignment (0.1 %, plus 0.3 min), mass alignment (5 ppm, plus 2 mDa) and minimum filter matches (features must be present in 66 % of the samples in at least one sample group). At the end, manual cleanup was performed by means of Find by Ion filters, which enabled keeping only features present in 25 % across all sample files.

In addition, features containing abnormal peak shapes or unusual isotopic distributions were discarded after checking all the features one by one again. 232 entities resulted from this data pretreatment process distributed through all retention times in the chromatogram (see Fig. 5.13).

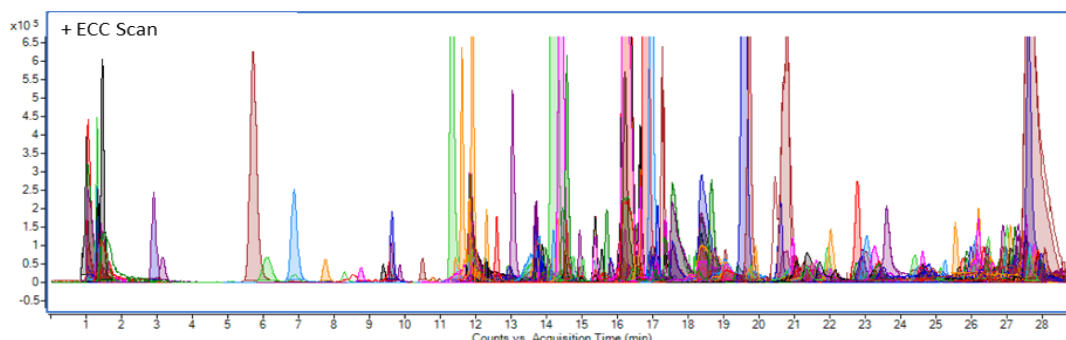


Fig. 5.13. An example of Extracted Compound Chromatogram (ECC) of the molecular features found in a plasma sample from a paediatric CKD patient.

Finally, the abundances of all the features obtained for each data file were summed and represented as a function of time (see Fig. 5.14). This could be considered somehow a data quality assessment to be carried out before great efforts are put on data analysis. Neither increase of the signal, that could be indicative of carry over or accumulation of metabolites in the system, nor decrease, that could be related with a loss of sensitivity due to signal suppression, were observed. Therefore, data were considered *a priori* suitable for data analysis.

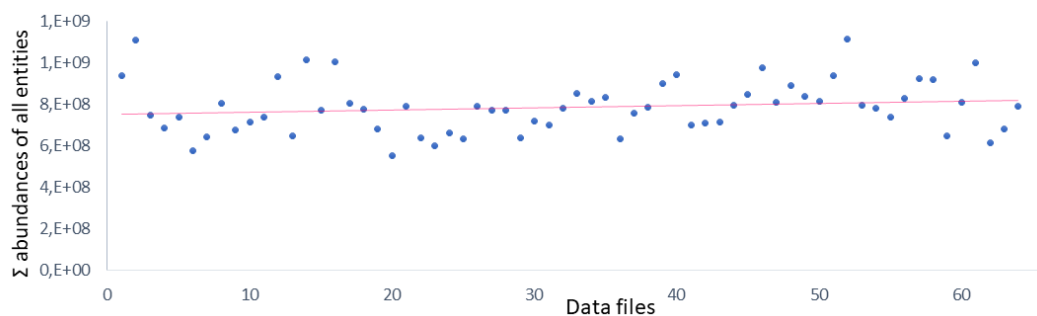


Fig. 5.14. Data files in the order followed by the batch represented opposite the sum of the abundances of the features obtained for each data file. A trend line is represented in pink colour.

5.6.5. DATA ANALYSIS

Data mining and data analysis are known to be time-consuming steps in untargeted metabolomics studies. For that reason, preliminary data analysis for data quality assessment is interesting to avoid laborious work and get unexpected results. Accordingly, PCA was used to check that all the QC samples analyzed during sample analyses gathered and did not present any particular trend. QC samples clustered together and are placed between control and CKD samples, which means that the system kept stable during the analysis and sample pretreatment performed alike, as it can be observed in Fig. 5.15.

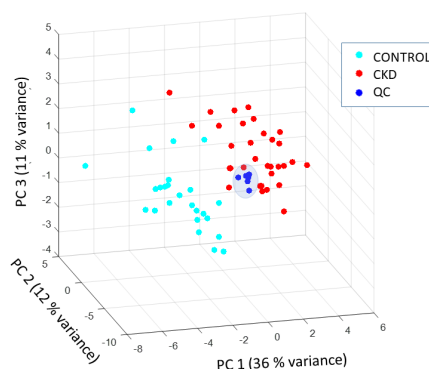


Fig. 5.15. PCA representation of control, CKD and QC samples.

5.6.5.1 Determination of statistically significant entities


It is common knowledge that the data analysis method used could affect the entities found to be relevant to discriminate between groups. Accordingly, two different chemometric methods and software were used and significantly up- or down-regulated entities matching in both chemometric approaches were considered exclusively, in order to assure with a high degree of confidence that the relevant features identified are not only a consequence of the data analysis method followed.

For that purpose, two different software were used: Mass Profiler Professional software (MPP) version B.13.00 (Agilent Technologies, Santa Clara, CA, USA) and Matlab R2015a (Mathworks, Natick, Massachusetts, USA). The data generated in Profinder software was exported to suitable format for each software. Comma-separated values files (.CSV file) containing detailed area were exported for later use in Matlab and compound exchange format files (.CEF file) for using them in MPP.

Data analysis was first carried out using **MPP software** following the workflow suggested by the software. It is of interest to note that MPP automatically performs logarithm

transformation with the purpose of complying with heteroscedasticity reduction. The conditions set for MPP software consisted on data scaling (Z-transform), filtering by flags (to find features present with acceptable values in the 33 % of the samples), filtering by frequency (entities that appear in at least 90 % of all samples in CKD or control group) and finally, a significance analysis (unpaired T test with asymptotic p -value computation, Benjamini-Hochberg multiple testing correction and p -value ≤ 0.05 , with a fold-change cut-off value ≥ 2). After applying this data analysis to data from CKD and control plasma sample analysis, the process returned 7 significant entities, as showed in Table 5.6. The notation used to represent the entities is the following: M (estimated with Profinder algorithm from experimental m/z) @ retention time (expressed in minutes).

Table 5.6. Summary of the results obtained from data analysis carried out in MPP software.



Compound (M@RT)	p -value	Corrected p -value	Absolute fold change	Up- or down regulation
125.0836@1.32	1.74×10^{-8}	8.91×10^{-7}	2.49	Up
231.1477@1.32	2.42×10^{-8}	9.94×10^{-7}	2.47	Up
246.233@15.61	1.37×10^{-8}	8.91×10^{-7}	2.50	Up
303.2036@2.94	2.13×10^{-5}	2.57×10^{-4}	2.07	Up
313.2242@11.68	1.91×10^{-5}	2.57×10^{-4}	2.08	Up
313.2242@12.15	1.34×10^{-7}	3.04×10^{-4}	2.38	Up
584.2613@12.46	2.82×10^{-12}	5.80×10^{-10}	2.88	Down

Besides, another data treatment procedure was followed using **Matlab software**, with the aim of comparing results from Matlab with the significant entities obtained with MPP software, to select only features matching in both data analysis methodologies. Logarithm transformation of data was carried out in order to obtain a more symmetric distribution. Afterwards, data was autoscaled in order to give the same statistical weight to all metabolites, regardless of their abundance in plasma. Then, Sparse Partial Least Squares Discriminant Analysis (SPLS-DA) was applied to reduce the high number of features.

SPLS-DA needs for optimization of the minimum number of entities to be selected. For that purpose, raw data was log transformed and divided into training (80 % of the data set) and test sets (20 % of the data set). Thereafter, autoscaling was performed and SPLS-DA was used to get sample group classification performance in the training set using leave-one-out (LOO) cross-validation with 10, 15, 20, 25, 30, 35, 40, 45 and 50 entities:

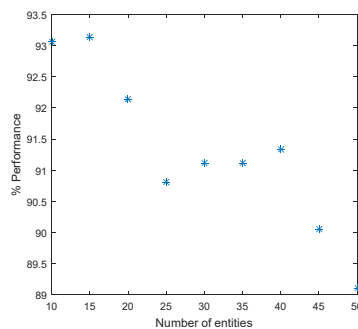


Fig. 5.16. Optimization of the number of entities to be selected by means of SPLS-DA using 10, 15, 20, 25, 30, 35, 40, 45 and 50 entities.

It has to be noted that the interest is towards finding the smallest amount of entities that returns maximum performance. The highest performance was obtained with 10-15 entities. Taking into account these results and with the aim of selecting the most appropriate number of entities during variable selection step, again the number of entities was optimized using SPLS-DA with entities ranging from 5 to 15 (see Fig. 5.17). Again, the objective was to find the minimum number of entities for which a significantly high performance was acquired. Figure 5.17 shows increasing performance when adding features until model built with 8 entities, where an elbow in the graph is observed. From 8 entities onwards performance is similar, thus the addition of new features does not contribute significantly to the model. Moreover, even though the highest performance is obtained for 13 entities, only 1.5% better performance is obtained with the addition of 5 entities. Besides, PCA analysis and subsequent representation showed that with entities from 8 entities upwards there was a considerable separation between CKD and control groups, and this separation did not improve when adding new features to the model. Finally, Student's *t*-test was used to check whether all the selected entities were statistically significant when comparing disease and control groups.

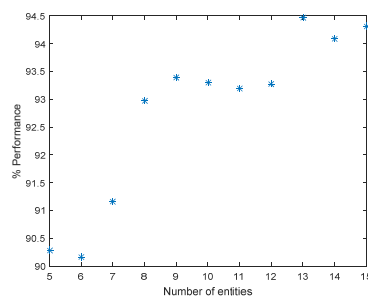



Fig. 5.17. Optimization of the number of entities by means of SPLS-DA using 5 to 15 entities.

After the optimization of the ideal number of entities, SPLS-DA model was built for 8 entities and the following features were found to be significantly altered according to the data analysis procedure carried out in Matlab.

Table 5.7. Summary of the results obtained from data analysis carried out in Matlab software.



Compound (M@RT)	Up- or down regulation
125.0836@1.32	Up
231.1477@1.32	Up
243.1460@1.40	Up
246.2333@15.61	Up
313.2242@12.15	Up
423.3344@20.05	Up
582.2473@13.69	Down
584.2613@12.46	Down

Due to the fact that data analysis method used is known to affect the significant entities found, it was decided to select only the features concurring in both chemometric approaches. Five out of the ten entities obtained from MPP and Matlab procedures matched in both approaches, as it can be observed in the Venn diagram from Fig. 5.18. Therefore, only these five features were studied as potential biomarkers.

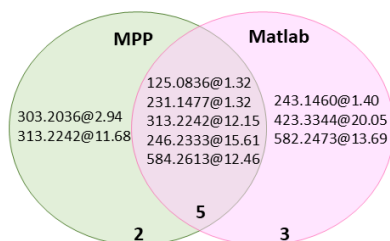


Fig. 5.18. Venn diagram showing the significant features obtained with MPP and Matlab.

Regarding that the goal of this study is to find compounds able to improve differentiation between control and CKD groups, it was consider important to verify that both peak shape and spectra of these five entities correspond to potential compounds, not to artifacts. For that purpose, EIC chromatograms for these five entities that are going to be studied as potential biomarkers in a random plasma sample were obtained and an example chromatogram can be observed in Fig. 5.19.

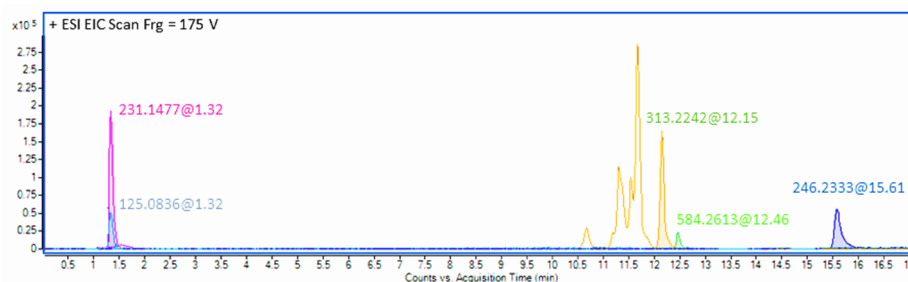


Fig. 5.19. EIC chromatogram of matching significant entities.

Once selected these significant features to discriminate between disease and control groups, PCA was carried out and samples were represented according to the new derived variables or principal components (PC), showing good separation between groups in addition to an accumulated variance of 82 % using 3 principal components. Besides, taking into account the loadings, PCA provides more information regarding relative concentration of these features in CKD and control groups. Feature 584.2613@12.16 is the only one out of the five entities decreased in CKD paediatrics, whereas the rest of the entities were found to be increased.

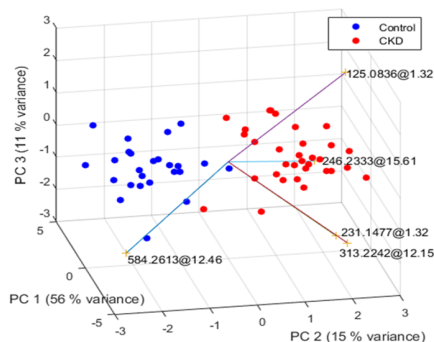


Fig. 5.20. PCA obtained using the five matching significant analytes showing differentiation between control and CKD sample groups.

Even though PCA is a very powerful explorative unsupervised method, the use of supervised classification methods, such as PLS-DA, is preferred. PLS-DA was used to obtain the resulting performance from the classification of samples into control or CKD groups according to the models built with these features with prior knowledge of sample grouping. Data set was divided into training (80 %) and test (20 %) sets and the optimization of the latent variables (LV) was carried out doing LOO cross-validation in the training set. In this case, the use of 1 LV provided the best result. Once selected the optimal number of latent variables, model was built using the training set and it was applied to the test set to obtain the performance of the model. The operation of dividing

data into training and test sets was repeated 50 times and the average performance was achieved. PLS-DA (LV=1) for the classification of control and CKD samples showed an overall performance of 96 %.

Similarly, PCA using only the five concurring entities was carried out and represented considering only early CKD patients, those at CKD 2nd stage, against controls. There is a separation between these groups, as it can be observed from Fig. 5.21. Besides, the distribution of the loading vectors is quite similar to the one using all the CKD patients regardless of the degree of the disease.

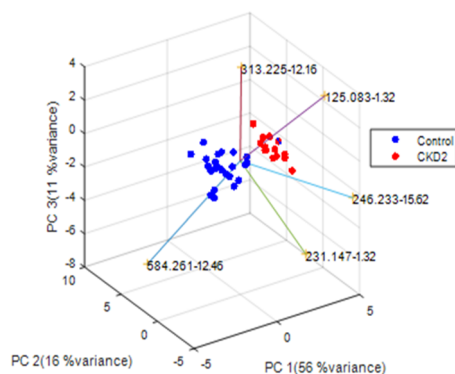


Fig. 5.21. PCA obtained using the five matching significant analytes showing the differentiation between control and early CKD patients.

In addition, when considering only early CKD patients (those at CKD 2 stage) against controls using PLS-DA (LV=1), a performance of 97 % was achieved.

Consequently, it could be affirmed that these matching relevant features could be interesting potential biomarkers for both general and early CKD.

5.6.5.2. Further data analysis on the significant entities

The relationship between the significant features and different conditions, such as, sex, age, treatment and CKD stage of the patients was also studied, as this information is frequently included in the equations used for CKD diagnosis and evaluation. For that purpose, PCA was performed with the abundance of these 5 entities and samples were represented coloured by the selected condition.

PCA coloured by **sex** showed that male and female samples were intermingled and no clear separation, relation or specific distribution between male and female paediatrics was found regarding these 5 features (see Fig. 5.22).

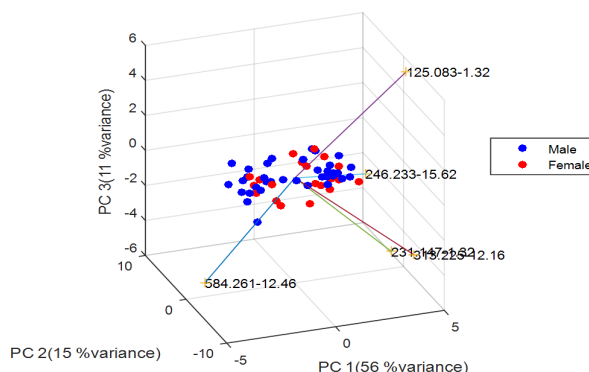


Fig. 5.22. PCA of significant entities coloured by sex.

Similarly, two different **age** groups were made to differentiate between children and adolescents, one with patients ranging from 2 to 12 year old and another one from 13 to 18 year old. PCA coloured according to these two groups did not show any clear distribution or separation between groups, as it can be interpreted from Fig. 5.23.

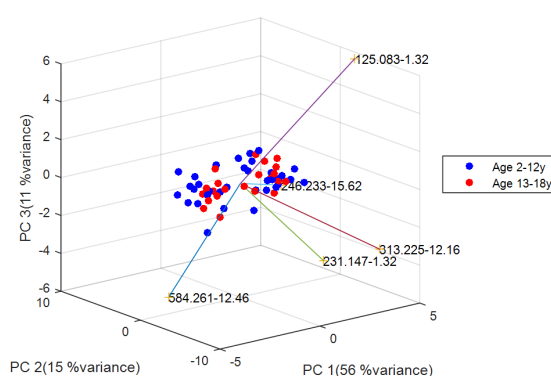


Fig. 5.23. PCA of significant features coloured by age groups.

Even though no clear separation between the two age groups was found, it was considered that some linear relation could exist between age and concentration of these entities. This was evaluated by means of patient age prediction with PLS regression, using LOO approach. This operation was performed for all samples together as well as differencing between control and for CKD samples. Having a look at the graphs in Figure 5.24, X-axis represents the true age of each patient, whereas Y-axis refers to the predicted age according to the PLS regression model built using the abundances of the 5 selected features. In case there was any linear correlation between feature concentration and age,

the prediction of the age would approximate to the real age. Nevertheless, the prediction was not good neither using all the samples together nor CKD samples alone, even though a slight trend could be perceived in control samples.

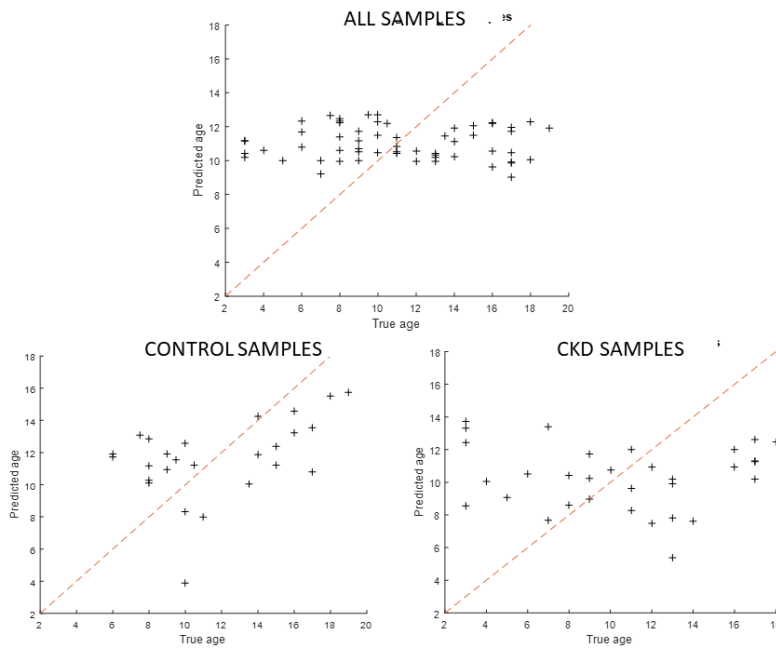


Fig. 5.24. Linear regression using the selected 5 entities for differentiated groups of samples.

To sum up, taking into account these two different multivariate approaches, no relation between age and feature concentration can be stated.

The effect of receiving or not **renal replacement therapy** (RRT), either dialysis or transplant, was evaluated by means of performing PCA on samples from paediatrics with CKD and colouring the graphical representation depending on the treatment received (see Fig. 5.25). Neither different distribution of the samples nor specific groupings were found in relation with the treatment received. Therefore, any defined trend on feature concentration according to the RRT received is not expected.

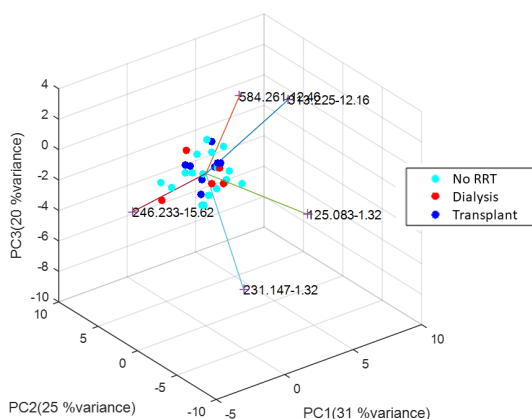


Fig. 5.25. PCA of CKD patients coloured according to the RRT received.

Furthermore, PCA was coloured according to the **CKD stage or degree** with the aim of finding any possible gradation according to feature concentration. Even though no clear separation of different CKD stages was found according to the abundance of these features, some trend was observed regarding the distance in the PCA of the control patients with the mildest CKD and the most severely ill patients, as showed in Fig. 5.26. For that reason, performing a PLS-DA sample classification was considered reasonable, following the previously explained sample division procedure for latent variable optimization and overall performance obtaining. PLS-DA sample classification (LV=1) lead to an overall performance of 47 % on average, which was considered low.

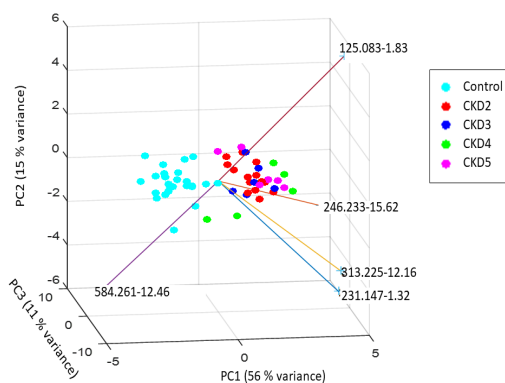


Fig. 5.26. PCA of control and CKD patients coloured according to the degree of the disease.

However, example cross-tables based on PLS-DA model obtained were represented to evaluate its performance and the results were better than expected, as it can be concluded from Table 5.8. Indeed, almost all the samples were correctly or one level above or below classified.

Table 5.8. An example of prediction of the degree of samples by means of PLS-DA (LV=1) model.

	CTRL _{pred}	CKD2 _{pred}	CKD3 _{pred}	CKD4 _{pred}	CKD5 _{pred}
CTRL _{known}	4	1	0	0	0
CKD2 _{known}	0	0	2	0	0
CKD3 _{known}	0	0	1	0	0
CKD4 _{known}	0	0	1	0	0
CKD5 _{known}	0	0	1	0	0

Therefore, it could be concluded that feature concentration might somehow be affected by the CKD degree of the patients. It has to be noted that a relation between CKD degree and feature concentration is desired when looking for potential biomarkers.

To sum up, plasma levels of these 5 features do not seem to be affected by sex, age and treatment received, whereas their concentration varies to some extent in accordance with the seriousness of the disease.

5.6.6. METABOLITE IDENTIFICATION

Metabolite identification aims at the determination of the chemical structure of the 5 statistically significant concurrent metabolites that have showed to be interesting to differentiate between control and CKD paediatrics. The first level at metabolite identification is based on the comparison of experimental accurate mass acquired in MS mode and theoretical exact mass in several databases. Once putative compounds have been established, MS/MS spectra of these features are obtained and compared with reference MS/MS spectra. Finally, the use of analytical standards enables identity confirmation of the metabolites by means of retention time, MS and MS/MS spectra comparison.

5.6.6.1 MS libraries

The use of the different databases such as METLIN, HMDB, LIPID MAPS and mzCloud allowed the putative identification of 3 out of 5 features (see table 5.9). It has to be noted that when comparing experimental and theoretical masses, a mass tolerance below 5 ppm was accepted.

The quality of compound identification is assessed using overall identification scores or ID scores⁵⁷. ID scores consider the confidence of compound identification according to the

accurate mass, isotope patterns, relative abundances, m/z distances and the error of the accurate mass measurement (expressed as parts per million). In this case, these values have been obtained from the Molecular Feature Extraction algorithm provided by MassHunter Qualitative Analysis software. ID scores are known to be affected by both signal intensity and matrix effect. For the putative identification of these features the overall ID scores were above 95 % in all the cases and the error of the accurate mass measurement below 3.87 ppm.

Table 5.9. Putative compound identification using the aforementioned libraries.

RT (min)	Experimental m/z	Ion species	Putative identification	Formula	Theoretical m/z	Δ ppm	ID Score (%)	Up- or down-regulated
1.32	126.0930	[M+H] ⁺	Not identified					Up
1.32	232.1552	[M+H] ⁺	<i>i</i> -butyrylcarnitine <i>n</i> -butyrylcarnitine	C ₁₁ H ₂₁ NO ₄	232.1543	3.87	95.57	Up
15.61	264.2685	[M+NH ₄] ⁺	Not identified					Up
12.15	314.2327	[M+H] ⁺	Cis-4-decenoylcarnitine 9-decenoylcarnitine	C ₁₇ H ₃₁ NO ₄	314.2326	0.31	98.13	Up
12.46	585.2722	[M+H] ⁺	Bilirubin	C ₃₃ H ₃₆ N ₄ O ₆	585.2708	2.39	98.39	Down

Profinder software estimates that these m/z are the protonated species of the metabolites, except for 264.2685 m/z for which [M+NH₄]⁺ adduct formation was suggested by the algorithm. However, taking into account that neither the mobile phase nor the reagents used for sample treatment contained ammonia, it was considered odd that an adduct with NH₄ could be formed. For that reason, the possibility that 264.2685 m/z could correspond to [M+H]⁺ adduct like the previous metabolites was studied, but this search did not return any candidate. Therefore, 264.2685 m/z was examined as a fragment ion in mzCloud library, which contains a tool to search MSⁿ ion products, just in case the significant entity was the product of fragmentation of another molecule. According to this search, 264.2685 m/z entity could correspond to a [M]⁺ fragment of sphingosine-1-phosphate (C₁₈H₃₄N). This metabolite could be involved in renal diseases according to previous research^{58,59}. With the aim of checking whether the entity could be fragmenting in the ionization source, plasma sample analysis was repeated under the same conditions of Table 5.5 but with a lower fragmentor voltage (Frg=100 V). This analysis confirmed that the fragment could correspond to sphingosine-1-phosphate, as the m/z matching with the protonated specie of this metabolite, 380.2560 m/z , appears at the same retention time and increases using a Frg=100 V instead of 175 V (see fig. 5.27).

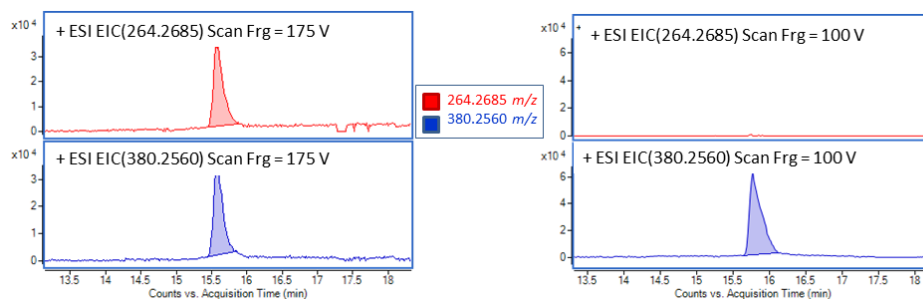


Fig. 5.27. EIC chromatograms for 264.2685 m/z and 380.2560 m/z obtained with a fragmentor voltage of 100 V and 175V.

Consequently, it was concluded that 264.2685 m/z could correspond to a product of fragmentation in the ionization source of sphingosine-1-phosphate when a fragmentor voltage of 175 V is used.

Table 5.10. Putative compound identification for 264.2685@15.61 min entity.

RT (min)	Experimental m/z	Ion species	Putative identification	Formula	Theoretical m/z	Δ ppm	ID Score (%)	Up- or down-regulated
15.61	380.2570	[M+H] ⁺	Sphingosine-1-phosphate	C ₁₈ H ₃₈ NO ₅ P	380.2560	1.88	97.09	Up

5.6.6.2. MS/MS fragmentation experiments

Putative identification of the features found to be significant in CKD and control paediatric class discrimination needs to be confirmed by means of characteristic MS/MS spectra. MS/MS fragmentation and isotopic distributions of precursor and product ions were studied and compared with MS/MS spectral data from the libraries listed in section 5.2.6. For that purpose, experiments were repeated in targeted MS/MS mode, in 30-1000 m/z mass range, with an acquisition rate of 1 spectra/s (in MS/MS mode), collision energies ranging from 10 V to 40 V and the analysis conditions showed in Table 5.5. Mass accuracy of the recorded ions was assured by means of continuous internal calibration with 121.0509 m/z and 922.0098 m/z .

Due to the fact that no candidates were found for **126.0930@1.32**, MS/MS fragmentation was performed with the aim of obtaining more information about this entity. MS/MS spectra with a collision energy of 30 V of 126.0930 m/z showed 4 characteristic fragments: 41.0395 m/z , 69.0306 m/z , 70.0666 m/z and 98.0597 m/z . These fragments were searched in mzCloud library and the peaks 41.0375 m/z , 70.0633 m/z and 98.0580 m/z were found in reference MS/MS spectra of 2-aminoadipic acid.

However, 126.0930 m/z did not match with the $[M+H]^+$ of 2-aminoadipic acid (162.0761 m/z). For that reason, the use of a lower fragmentor (Frg=100 V) was assayed to verify that no in-source fragmentation was occurring at that voltage. Nevertheless, $[M+H]^+$ corresponding to 2-aminoadipic acid was not found either using Frg=100 V or Frg=175 V, so that the analysis of an analytical standard was required in order to verify if 126.0930@1.32 could correspond to 2-aminoadipic acid.

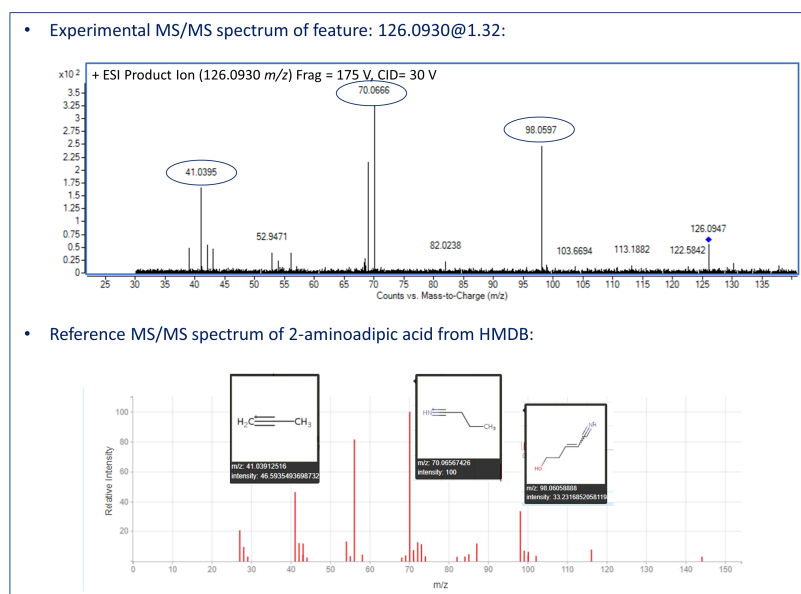


Fig. 5.28. Experimental spectrum of 126.0930@1.32 entity in comparison with MS/MS reference spectrum of 2-aminoadipic acid.

Regarding **232.1552@1.32** feature, MS/MS analysis was carried out in order to verify if this entity could correspond to a carnitine and, if possible, to discern between the two carnitine candidates: *i*-butyrylcarnitine and *n*-butyrylcarnitine. MS/MS analysis of 232.1552 m/z with a collision energy of 30 V showed a base peak close to 85.0295 m/z (see Fig. 5.29), which corresponds to the fragment $C_4H_5O_2$, a characteristic peak in MS/MS spectra of all carnitines⁶⁰. In addition, some other minor peaks were found in MS/MS spectrum but were not conclusive for identification purpose. Thus, this MS/MS spectrum might correspond to a carnitine and is compatible either with *i*-butyrylcarnitine or *n*-butyrylcarnitine, even though it is not possible to differentiate between them. Analytical standard analysis of both candidates is necessary to establish the identity of 232.1552@1.32.

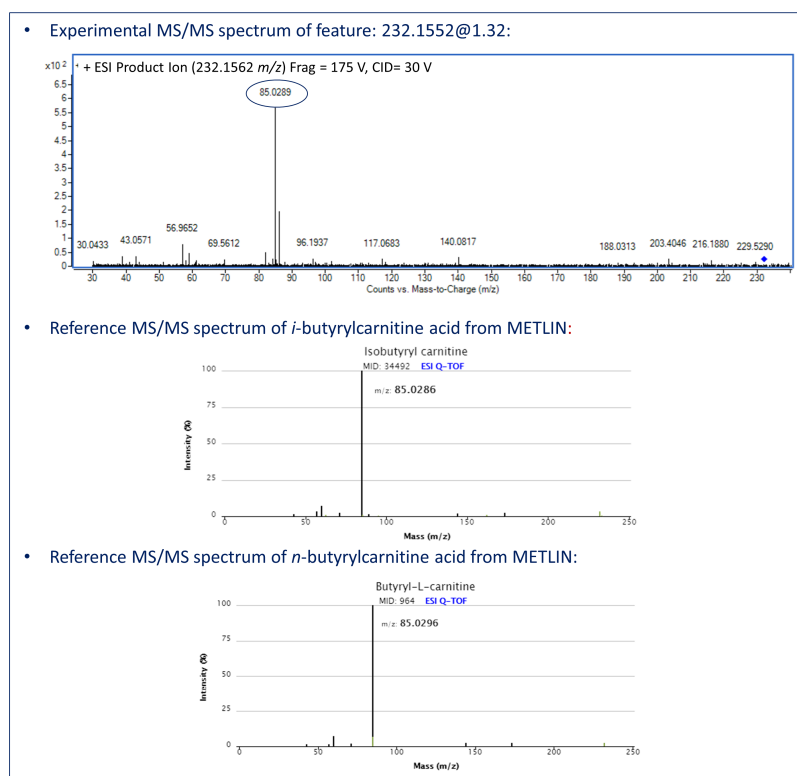


Fig. 5.29. Experimental MS/MS spectrum of 232.1552@1.32 compared to reference MS/MS spectra of *i*-butyrylcarnitine and *n*-butyrylcarnitine.

Similar to the previous entity, **314.2322@12.15** was suspicious of being a carnitine, so that a characteristic base peak close to 85.0295 m/z was expected. MS/MS spectra from plasma samples with a collision energy of 30 V showed that there was an indicative base peak around 85.0295 m/z , thus confirming that the feature may correspond to a carnitine. According to different libraries, the two candidates proposed for this m/z were carnitines, so this spectrum could correspond to either of them: *cis*-4-decenoylcarnitine and 9-decenoylcarnitine. The comparison of MS/MS spectra of both candidates could clarify the identity of the feature. Nevertheless, no reference spectra was found for *cis*-4-decenoylcarnitine and 9-decenoylcarnitine. A study from Giesbertz et al. showed that the MRM transition used to monitor decenoylcarnitine, without distinction of the isomer, was 370.2>85.1. Therefore, in this case the use of analytical standards is indispensable to identify correctly this feature.

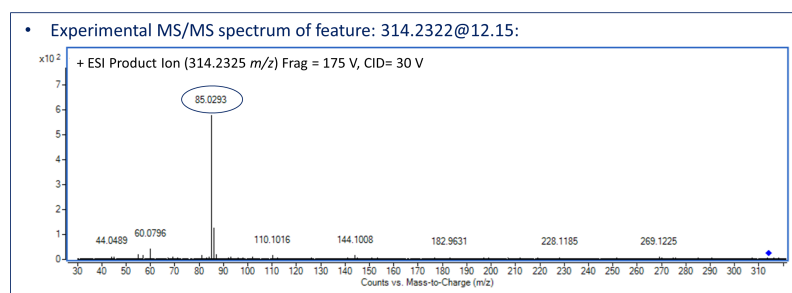


Fig. 5.30. Experimental MS/MS spectrum of 314.2322@12.15.

The fragmentation pattern of **585.2722@12.46** was studied using a collision energy of 30 V. A characteristic fragment close to 299.1396 m/z was found, which according to reference MS/MS spectra from METLIN library, could correspond to a cleavage of the central C-C of bilirubin. As a consequence, the analysis of bilirubin analytical standard may confirm the identification of this feature as bilirubin.

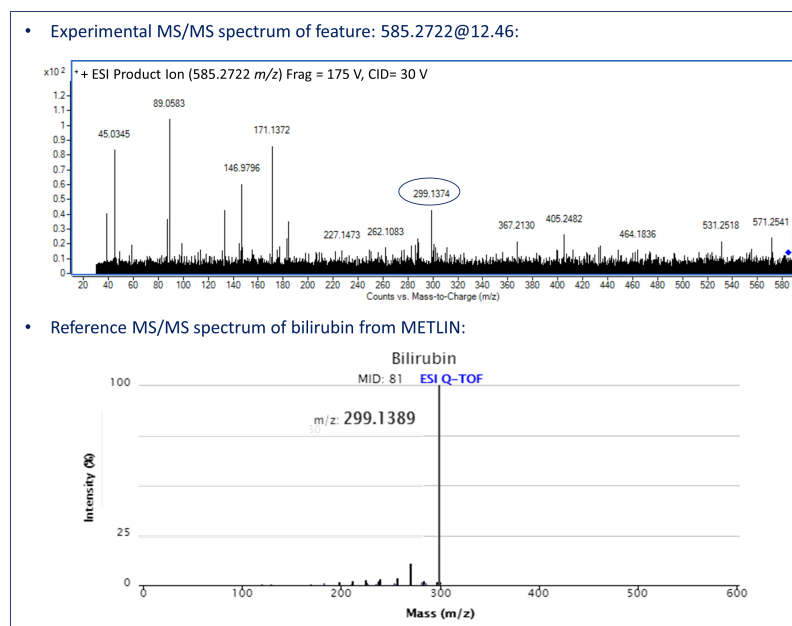


Fig. 5.31. Experimental MS/MS spectrum of the feature 585.2722@12.46 and a reference MS/MS spectrum of bilirubin extracted from METLIN database.

Finally, MS/MS fragmentation patterns of both **264.2685@15.61**, the hypothetical fragment of sphingosine-1-phosphate with a fragmentor voltage of 175 V (collision energy 30 V) and 380.2560@15.61, the theoretically entire molecule of the same metabolite with a fragmentor voltage of 100 V (collision energy 40V) were examined. As it can be observed in Fig. 5.32, MS/MS spectra of 380.2560 m/z contained the peak 264.2685 m/z , which could correspond to the tail of the sphingosine-1-phosphate ($C_{18}H_{34}N$), matching with

The analysis of 2-DL-aminoadipic acid analytical standard in MS mode showed a peak at retention time 0.92 min, which did not match with the retention time of the feature of interest in plasma, **126.0930@1.32** (Fig. 5.33). Moreover, 116.1072 m/z and 144.0658 m/z were the two major ions in the spectrum of the analytical standard, in addition to less abundant signal 162.0761 m/z ($[M+H]^+$), and did not contain the 126.0930 m/z of the feature. Besides, EIC chromatograms of 126.0930 m/z were compared in plasma samples before and after spiking 2-DL-aminoadipic acid, just in case this metabolite may ionize differently in plasma matrix. However, the addition of the analytical standard did not show any increase of this m/z , so that this possibility was discarded.

To sum up, it could be stated that 126.0930 m/z does not correspond to a fragment of the aforementioned metabolite as 2-DL-aminoadipic acid eluted 0.4 min before the feature of interest and MS spectra did not match. In addition, MS/MS fragments from the feature in plasma matching with reference spectra might correspond to common and inespecific fragmentations. Therefore, it could be concluded that 126.0930 m/z may correspond to a metabolite with some physicochemical properties similar to 2-DL-aminoadipic acid, but not exactly to that compound.

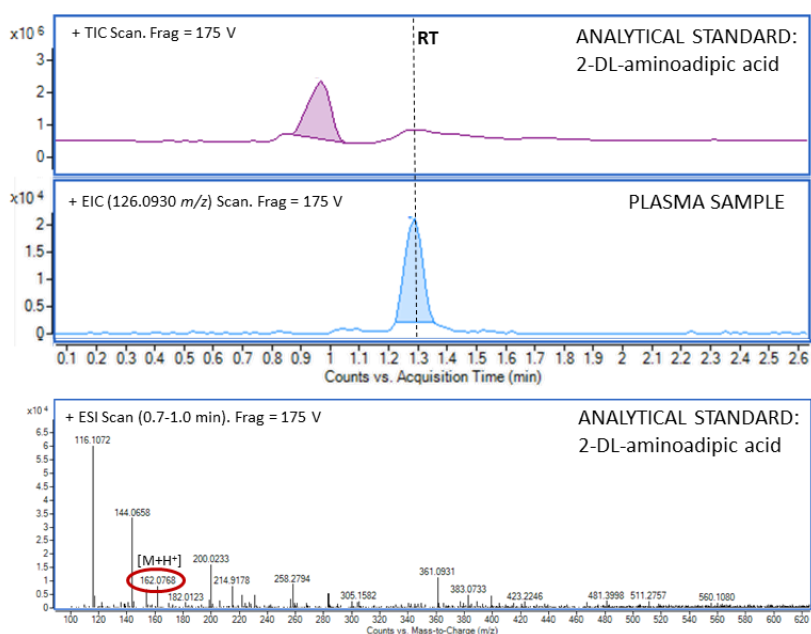


Fig. 5.33. TIC scan and MS spectrum of 2-DL-aminoadipic acid standard in comparison with EIC of the 126.0930@1.32 feature in plasma.

Concerning the identification of **232.1552@1.32** feature, both *i*-butyrylcarnitine and *n*-butyrylcarnitine analytical standards were analyzed in MS mode. Even though the same MS spectra was obtained for these analytical standards, the retention times were different. The retention time of *n*-butyrylcarnitine matched with the retention time of 232.1552@1.32 entity in plasma (see Fig. 5.34). Besides, MS/MS spectra of *n*-butyrylcarnitine was studied and the profile coincided with the MS/MS fragmentation pattern of this feature in plasma. Consequently, the potential biomarker was identified as *n*-butyrylcarnitine.

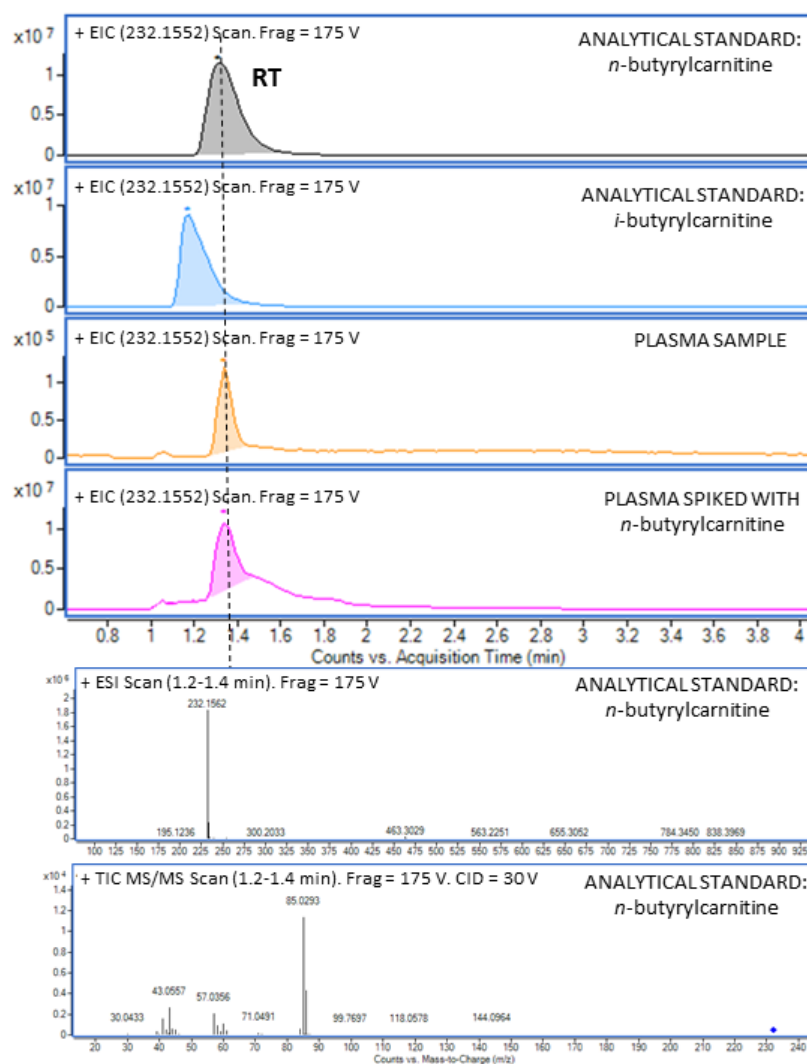


Fig. 5.34. EIC of the analytical standards *n*-butyrylcarnitine and *i*-butyrylcarnitine are compared with 232.1552@1.32 feature, showing matching retention times between the feature and *n*-butyrylcarnitine. The addition of *n*-butyrylcarnitine standard to plasma sample is observed as well. Below, MS and MS/MS spectra of *n*-butyrylcarnitine are showed.

Regarding **314.2322@12.15** feature, the availability of analytical standards *cis*-4-decenoylcarnitine and 9-decenoylcarnitine to verify its identity was more limited and none of the two candidates were commercialized. Therefore, *cis*-4-decenoylcarnitine was synthesized on-demand by an organic chemistry research group at Universidad Autónoma de Madrid (Madrid, Spain). This experiment led to the same retention time and MS spectra for *cis*-4-decenoylcarnitine and 314.2322@12.15 feature. Moreover, when spiking plasma samples with this analytical standard, there was a significant increase of the chromatographic peak. Taking into account that both retention time and MS spectra matched, MS/MS analysis of *cis*-4-decenoylcarnitine standard was performed. MS/MS spectra of the analytical standard also coincided with the fragmentation pattern of this feature in plasma. Therefore, 314.2322@12.15 feature was identified as *cis*-4-decenoylcarnitine.

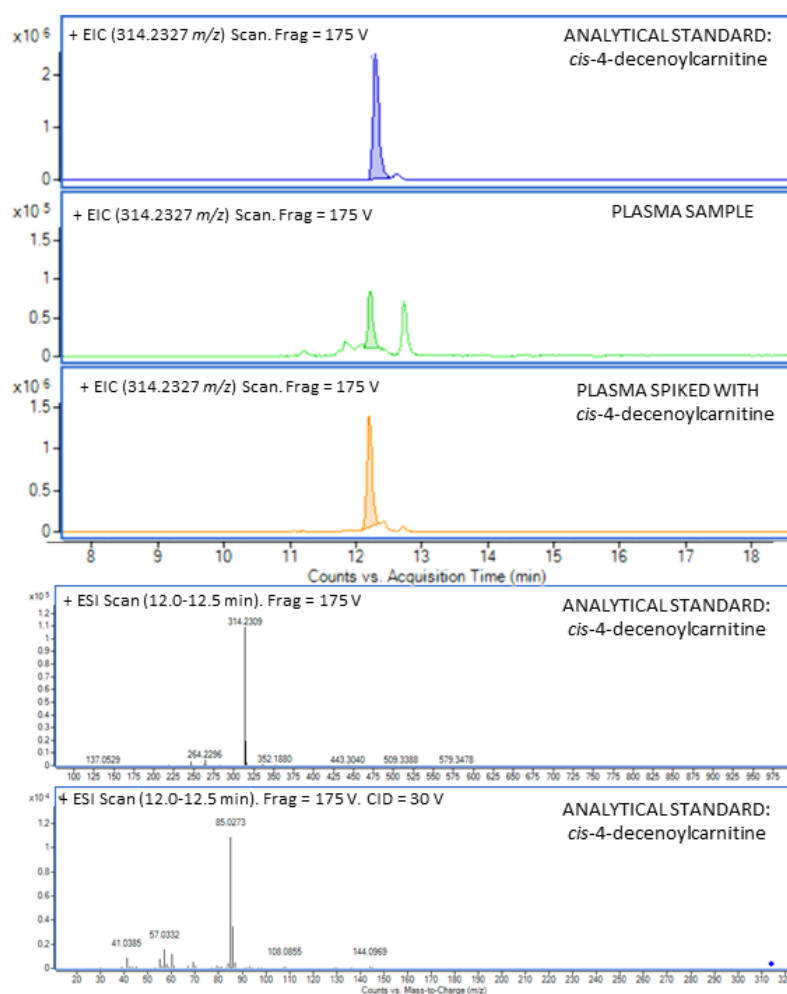


Fig. 5.35. EIC chromatograms of *cis*-4-decenoylcarnitine standard in comparison with EIC of 314.2327@12.15 in plasma. MS and MS/MS spectra are also showed for *cis*-4-decenoylcarnitine.

Besides, **585.2722@12.46** feature was compared with its candidate bilirubin analytical standard. For that purpose, bilirubin analytical standard and plasma were analyzed, obtaining a minimal shift in retention times. In addition, MS spectra from bilirubin matched perfectly with the m/z of the feature of interest in plasma. Moreover, when spiking plasma samples with bilirubin, an increase in EIC of the feature was observed (Fig. 5.36). Finally, MS/MS spectra of bilirubin standard was obtained and it was verified that both plasma sample and bilirubin had the same fragmentation pattern.

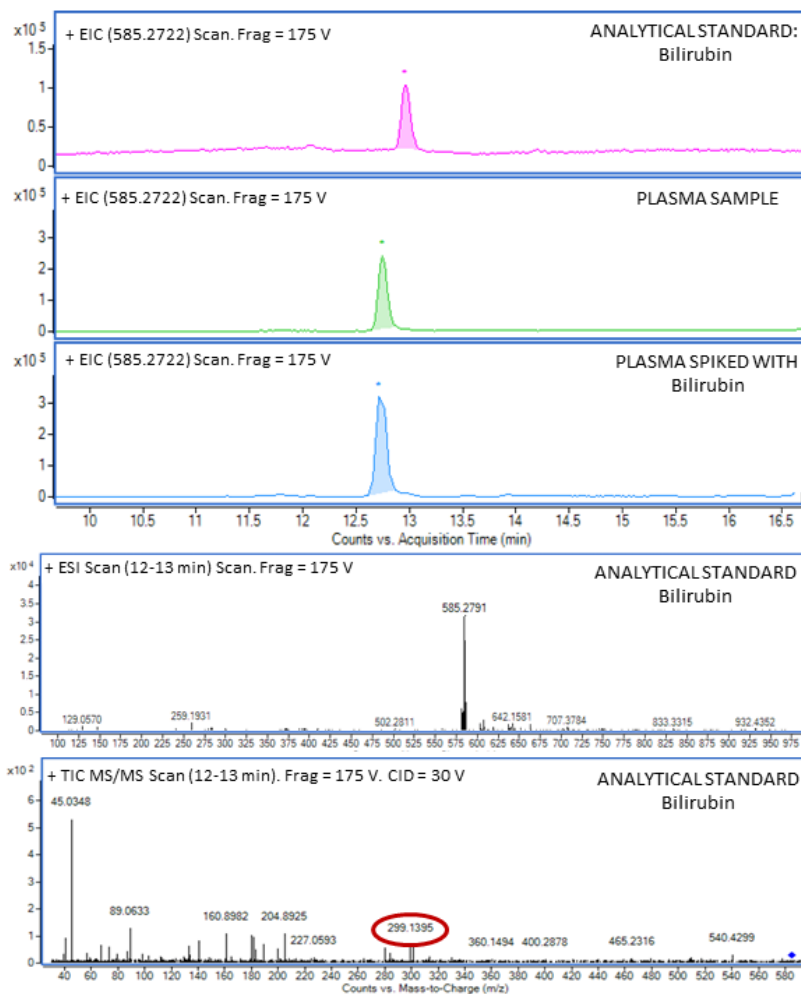


Fig. 5.36. EIC chromatograms of the analytical standard bilirubin in comparison with 585.2722@12.46 in plasma sample, before and after spiking it. In addition, an increase in the signal is observed when spiking plasma with bilirubin. MS and MS/MS spectra for bilirubin are also showed.

Regarding **264.2685@15.61** feature, sphingosine-1-phosphate analytical standard and plasma samples were analyzed in MS mode using 100 V fragmentor voltage and 380.2560 m/z was monitored. Retention time matched in plasma sample and analytical standard.

Moreover, the addition of this analytical standard to plasma samples leads to an increase of the EIC peak corresponding to the feature. Finally, MS/MS fragmentation pattern of 380.2560 m/z precursor ion of the analytical standard sphingosine-1-phosphate was obtained. A fragment of 264.2713 m/z was achieved from sphingosine-1-phosphate standard analysis, close to the 264.2685 m/z found to be significant in plasma when using 175 V fragmentor voltage, instead of 100 V.

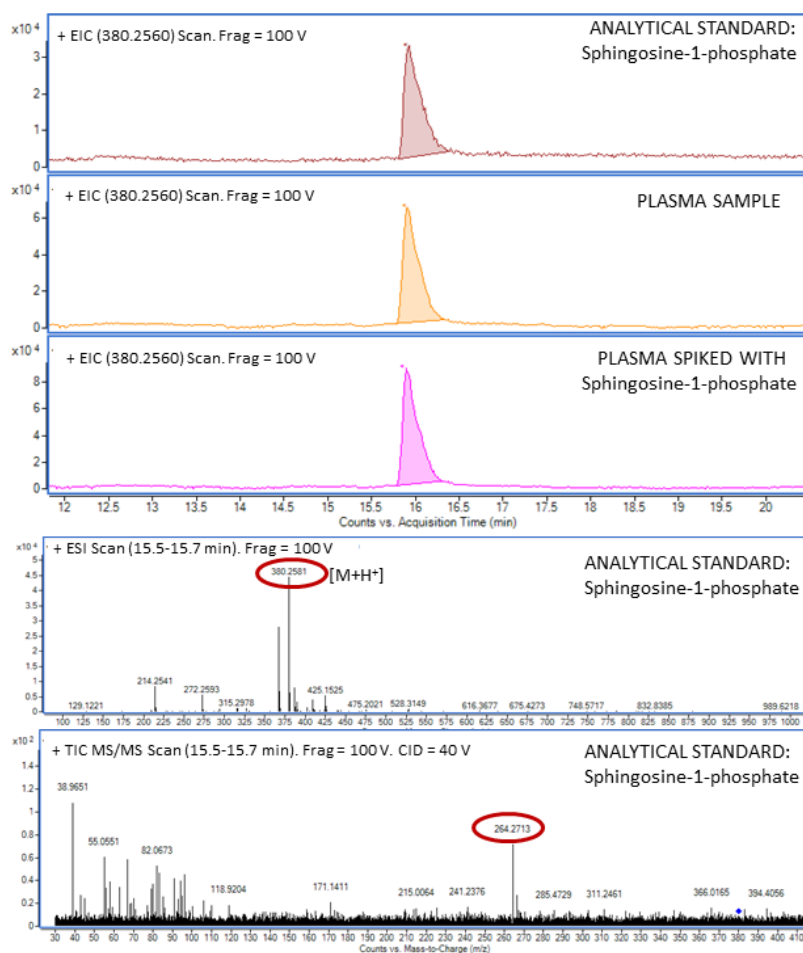


Fig. 5.37. EIC chromatograms of the analytical standard sphingosine-1-phosphate in comparison with EIC of 380.2560@15.61 in unspiked and spiked plasma sample. In addition, MS and MS/MS spectra of sphingosine-1-phosphate are showed.

The results of the identification and confidence levels of the significant entities to differentiate between control and CKD paediatrics, according to Metabolomics Standard Initiative³⁶ explained in section 5.2.6, are showed in Table 5.11. Four out of five compounds were successfully identified with the maximum confidence as *n*-butyrylcarnitine, sphingosine-1-phosphate, *cis*-4-decenoylcarnitine and bilirubin. However,

one feature was not successfully identified even though its retention time, MS and MS/MS spectra were obtained for future characterization.

Table 5.11. Identification of the significant features and identification confidence level, according to the Metabolomics Standards Initiative (MSI)³⁶.

RT (min)	Experimental m/z	Ion species	Identification	MSI Identification confidence level
1.32	126.0930	[M+H] ⁺	Not identified	4
1.32	232.1552	[M+H] ⁺	<i>n</i> -butyrylcarnitine	1
15.61	380.2570	[M+H] ⁺	Sphingosine-1-phosphate	1
12.15	314.2327	[M+H] ⁺	Cis-4-decenoyl carnitine	1
12.46	585.2722	[M+H] ⁺	Bilirubin	1

5.7. BIOLOGICAL INTERPRETATIONS OF THE RESULTS

The establishment of a relationship of these up- or down-regulated metabolites (see Fig. 5.38) with the disturbed metabolic pathways is of great interest to understand the causes and effects of impaired kidneys, as well as to find new potential therapeutic targets.

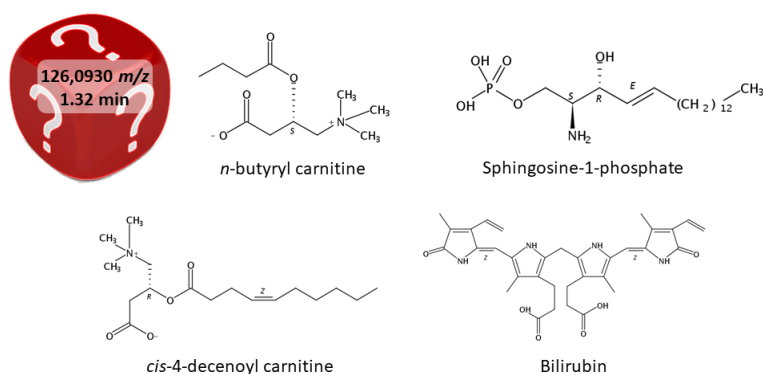


Fig. 5.38. Identified up- or down-regulated metabolites.

Regarding, *n*-butyrylcarnitine and *cis*-4-decenoylcarnitine, adult patients with ESRD on long-term hemodialysis are known to have free carnitine deficiency in addition to elevated acylcarnitine concentrations in plasma⁶¹. In fact, carnitine acts as an essential intermediary metabolite in lipid metabolism and complies with several functions, such as, shuttling acylcarnitines into mitochondria for beta-oxidation and energy production in fatty acid oxidation dependent tissues (e.g. cardiac and skeletal muscle)⁶². Moreover, the accumulation of acylcarnitines in plasma is known to be a consequence of decreased renal

clearance of the esterified carnitine fraction in patients not yet undergoing haemodialysis⁶³. Furthermore, increased butyrylcarnitine levels in plasma (with no distinction of the isomer) had already been found to be increased in adults with ESRD⁶¹. However, it is the first time, as far as we know, that *n*-butyrylcarnitine has been related with paediatrics with CKD. *Cis*-4-decenoyl carnitine has accordingly been proposed as potential biomarker of CKD in paediatrics for the first time. This metabolite had already been related to newborns having suffered from hypoxic ischemic encephalopathy, which presented altered levels in umbilical cord blood⁶⁴.

This untargeted metabolomics study would demonstrate that in addition to the lipids of the carnitine family, sphingolipids are also affected in paediatrics suffering from CKD. Sphingosine-1-phosphate is a signaling sphingolipid recognized as a decisive regulator of several metabolic pathways and has a role in several pathophysiological processes, like atherosclerosis, cancer, diabetes and osteoporosis⁶⁵. According to recent research, sphingosine-1-phosphate 1 receptor (S1P1R) activation blocks kidney inflammation and thus could relieve renal injury in rats with early-stage diabetic nephropathy. For that reason, targeting S1P1R has been suggested as an interesting therapeutic measure in diabetic nephropathy⁶⁶. In addition, some other studies have proven the expression of sphingosine kinases and S1P receptors in the kidney⁵⁸. Furthermore, sphingosine-1-phosphate formation is known to be stimulated by TGF- β 1 release, a protein which has also been suggested as a biomarker for CKD⁶⁷, thus leading to S1P1R activation⁶⁸.

Concerning the up-regulated feature with a mass-to-charge ratio of 126.0930 *m/z*, it could not be identified. However, considering its high polarity and MS/MS spectra, it is suspected that this feature could be some kind of amino acid derivative similar to amino adipic acid, which has been related to protein carbonyl oxidation in adults suffering from renal failure⁶⁹. Indeed, some amino acids and amino acid derivatives have been found to be increased in paediatrics with CKD⁷⁰⁻⁷².

Generally, an accumulation of metabolites is expected in CKD because of the nature of the disease. Nevertheless, according to these results paediatrics with CKD would show decreased bilirubin levels in plasma. These results match with recent studies performed in adults in which lower serum bilirubin concentration had already been associated with adverse renal outcomes⁷³. Moreover, serum total bilirubin concentration has been proposed as a novel risk factor for CKD progression in general population⁷⁴ and could help predicting progression of the disease in patients with moderate to severe CKD⁷³. Indeed, moderately elevated bilirubin levels could act as a protecting factor by lowering

progression of chronic kidney disease and related mortality in rats in connection with its antioxidant properties, protecting kidneys from circulating oxidative stress⁷⁵.

5.8. **CONCLUSIONS**

Taking into account the proposed objectives and the results obtained in this untargeted metabolomics study, it can be concluded that:

- ✓ A new LC-QTOF-MS analytical methodology has been optimized to extract as many features as possible and obtain the most relevant metabolites to differentiate between control and CKD paediatrics.
- ✓ The use of two different chemometric approaches showed that a high percentage of the significant features matched, showing 5 features that were significantly up- or down-regulated and enabled discrimination between groups.
- ✓ An overall accuracy of 96 % has been achieved for the classification of control and CKD samples using these 5 features (PLS-DA, LV=1), whereas considering only early CKD patients against controls a performance of 97 % was obtained (PLS-DA, LV =1). Consequently, it could be affirmed that these features could be interesting potential biomarkers for general and early CKD.
- ✓ The concentrations of the metabolites found to be significant are unlikely to be affected by age, sex or treatment received. On the contrary, some correlation have been found between metabolite concentration and CKD degree.
- ✓ The identification of the entities has been quite successful as 4 of these entities were fully identified as *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine, bilirubin and sphingosine-1-phosphate, with 1st level compound identification confidence according to the Metabolomics Standards Initiative. Furthermore, the only unidentified feature (and thus complying with 4th level from the MSI scale) was at least characterized by MS/MS spectra.
- ✓ Future development of a quantification methodology is needed for method validation, to check whether these promising metabolites might be suitable biomarkers, and for routine analysis of these metabolites in plasma. This will be covered in Chapter 6.

5.9. **BIBLIOGRAPHY**

1. Zhao YY. Metabolomics in chronic kidney disease. *Clin Chim Acta*. 2013; 422: 59-69.
2. Rhee EP. Metabolomics and renal disease. *Curr Opin Nephrol Hypertens*. 2015; 24: 371-379.
3. Patti GJ, Yanes O, Siuzdak G. Innovation Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*. 2012; 13: 263-269.
4. Theodoridis G, Gika HG, Wilson ID. Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. *Mass Spectrom Rev*. 2011; 30: 884-906.
5. Becker S, Kortz L, Helmschrodt C, Thiery J, Ceglarek U. LC-MS-based metabolomics in the clinical laboratory. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2012; 883-884: 68-75.
6. Zhang A, Sun H, Wang P, Han Y, Wang X. Modern analytical techniques in metabolomics analysis. *Analyst*. 2012; 137: 293-300.
7. Menon V, Sarnak MJ. The epidemiology of chronic kidney disease stages 1 to 4 and cardiovascular disease: a high-risk combination. *Am J Kidney Dis*. 2005; 45: 223-232.
8. Courant F, Antignac J-P, Dervilly-Pinel G, Le Bizec B. Basics of mass spectrometry based metabolomics. *Proteomics*. 2014; 14: 2369-2388.
9. Papadopoulos T, Bascands J-L, Schanstra JP, Klein J, Krochmal M, Cisek K, et al. Omics databases on kidney disease: where they can be found and how to benefit from them. *Clin Kidney J*. 2016; 9: 343-352.
10. Vuckovic D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal Bioanal Chem*. 2012; 403: 1523-1548.
11. Yin P, Lehmann R, Xu G. Effects of pre-analytical processes on blood samples used in metabolomics studies. *Anal Bioanal Chem*. 2015; 407: 4879-4892.
12. Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, et al. The human serum metabolome. *PLoS One*. 2011; 6: e16957.
13. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin Chem*. 2013; 59: 833-845.
14. Wedge DC, Allwood JW, Dunn W, Vaughan AA, Simpson K, Brown M, et al. Is Serum or Plasma More Appropriate for Intersubject Comparisons in Metabolomic Studies? An Assessment in Patients with Small-Cell Lung Cancer. *Anal Chem*. 2011; 83: 6689-6697.
15. Denery JR, Nunes AAK, Dickerson TJ. Characterization of Differences between Blood Sample Matrices in Untargeted Metabolomics. *Anal Chem*. 2011; 83: 1040-1047.

16. Brauer R, Leichtle AB, Fiedler GM, Thiery J, Ceglarek U. Preanalytical standardization of amino acid and acylcarnitine metabolite profiling in human blood using tandem mass spectrometry. *Metabolomics*. 2011; 7: 344-352.
17. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc*. 2011; 6: 1060-1083.
18. Gika H, Theodoridis G. Sample preparation prior to the LC--MS-based metabolomics/metabonomics of blood-derived samples. *Bioanalysis*. 2011; 3: 1647-1661.
19. Chen Y, Xu J, Zhang R, Abliz Z. Methods used to increase the comprehensive coverage of urinary and plasma metabolomes by MS. *Bioanalysis*. 2016; 8: 981-997.
20. Álvarez-Sánchez B, Priego-Capote F, Luque de Castro MD. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *TrAC-Trend Anal Chem*. 2010; 29: 111-119.
21. Alvarez-Sanchez B, Priego-Capote F, Luque dCMD. Metabolomics analysis II. Preparation of biological samples prior to detection. *TrAC Trends in Analytical Chemistry*. 2010; 29: 120-127.
22. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. *TrAC-Trend Anal Chem*. 2005; 24: 285-294.
23. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc*. 2010; 5: 1005-1018.
24. Issaq HJ, Veenstra TD. Proteomic and Metabolomic approaches to Biomarker Discovery. 2013.
25. Hyotylainen T. Sample collection, storage and preparation. *Chromatographic Methods in Metabolomics*. 18: Royal Society of Chemistry; 2013; p. 11-42.
26. Polson C, Sarkar P, Incedon B, Raguvaran V, Grant R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci*. 2003; 785: 263-275.
27. Bruce SJ, Tavazzi I, Parisod V, Rezzi S, Kochhar S, Guy PA. Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. *Anal Chem* 2009; 81: 3285-3296.
28. Want EJ, O'Maille G, Smith CA, Brandon TR, Uritboonthai W, Qin C, et al. Solvent-Dependent Metabolite Distribution, Clustering, and Protein Extraction for Serum Profiling with Mass Spectrometry. *Anal Chem*. 2006; 78: 743-752.
29. Michopoulos F, Lai L, Gika H, Theodoridis G, Wilson I. UPLC-MS-Based Analysis of Human Plasma for Metabonomics Using Solvent Precipitation or Solid Phase Extraction. *J Proteome Res*. 2009: 2114-2121.

30. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *J Pharm Biomed Anal.* 2014; 87: 12-25.
31. Lu X, Zhao X, Bai C, Zhao C, Lu G, Xu G. LC-MS-based metabonomics analysis. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2008; 866: 64-76.
32. Theodoridis GA, Gika HG, Want EJ, Wilson ID. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Anal Chim Acta.* 2012; 711: 7-16.
33. Le Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics.* 2011; 12: 253.
34. Le Cao KA, Rossouw D, Robert-Granie C, Besse P. A sparse PLS for variable selection when integrating omics data. *Stat Appl Genet Mol Biol.* 2008; 7.
35. Werner E, Heilier J-F, Ducruix C, Ezan E, Junot C, Tabet J-C. Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2008; 871: 143-163.
36. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics.* 2007; 3: 211-221.
37. Yu B, Zheng Y, Nettleton JA, Alexander D, Coresh J, Boerwinkle E. Serum metabolomic profiling and incident CKD among African Americans. *Clin J Am Soc Nephrol.* 2014; 9: 1410-1417.
38. Shah VO, Townsend RR, Feldman HI, Pappan KL, Kensicki E, Vander JDL. Plasma metabolomic profiles in different stages of CKD. *Clin J Am Soc Nephrol.* 2013; 8: 363-370.
39. Mutsaers HAM, Engelke UFH, Wilmer MJG, Wetzels JFM, Wevers RA, Van den Heuvel LP, et al. Optimized metabolomic approach to identify uremic solutes in plasma of stage 3-4 chronic kidney disease patients. *PLoS One.* 2013; 8: e71199.
40. Sharma K, Karl B, Mathew AV, Gangoiti JA, Wassel CL, Saito R, et al. Metabolomics reveals signature of mitochondrial dysfunction in diabetic kidney disease. *J Am Soc Nephrol.* 2013; 24: 1901-1912.
41. Atzori L, Mussap M, Noto A, Barberini L, Puddu M, Coni E, et al. Clinical metabolomics and urinary NGAL for the early prediction of chronic kidney disease in healthy adults born ELBW. *J Matern-Fetal Neonat Med.* 2011; 24: 41-44.
42. Toyohara T, Akiyama Y, Suzuki T, Takeuchi Y, Mishima E, Tanemoto M, et al. Metabolomic profiling of uremic solutes in CKD patients. *Hypertens Res.* 2010; 33: 944-952.
43. Hanna MH, Segar JL, Teesch LM, Kasper DC, Schaefer FS, Brophy PD. Urinary metabolomic markers of aminoglycoside nephrotoxicity in newborn rats. *Pediatr Res.* 2013; 73: 585-591.

44. Atzori L, Antonucci R, Barberini L, Locci E, Marincola FC, Scano P, et al. 1H NMR-based metabolic profiling of urine from children with nephrouropathies. *Front Biosci, Elite Ed.* 2010; E2: 725-732.
45. Beger RD, Holland RD, Sun J, Schnackenberg LK, Moore PC, Dent CL, et al. Metabonomics of acute kidney injury in children after cardiac surgery. *Pediatr Nephrol.* 2008; 23: 977-984.
46. Zivkovic AM, Yang J, Georgi K, Hegedus C, Nording ML, O'Sullivan A, et al. Serum oxylipin profiles in IgA nephropathy patients reflect kidney functional alterations. *Metabolomics.* 2012; 8: 1102-1113.
47. Sato E, Kohno M, Yamamoto M, Fujisawa T, Fujiwara K, Tanaka N. Metabolomic analysis of human plasma from haemodialysis patients. *Eur J Clin Invest.* 2011; 41: 241-255.
48. Posada-Ayala M, Zubiri I, Martin-Lorenzo M, Sanz-Maroto A, Molero D, Gonzalez-Calero L, et al. Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int.* 2014; 85: 103-111.
49. Godfrey AR, Williams CM, Dudley E, Newton RP, Willshaw P, Mikhail A, et al. Investigation of uremic analytes in hemodialysate and their structural elucidation from accurate mass maps generated by a multi-dimensional liquid chromatography/mass spectrometry approach. *Rapid Commun Mass Spectrom.* 2009; 23: 3194-3204.
50. Stec DF, Wang S, Stothers C, Avance J, Denson D, Harris R, et al. Alterations of urinary metabolite profile in model diabetic nephropathy. *Biochem Biophys Res Commun.* 2015; 456: 610-614.
51. Andrade F, Sanchez-Ortega A, Llarena M, Pinar-Sueiro S, Galdos M, Goicolea MA, et al. Metabolomics in non-arteritic anterior ischemic optic neuropathy patients by liquid chromatography-quadrupole time-of-flight mass spectrometry. *Metabolomics.* 2015; 11: 468-476.
52. Ciborowski M, Javier Ruperez F, Martinez-Alcazar MP, Angulo S, Radziwon P, Olszanski R, et al. Metabolomic Approach with LC-MS Reveals Significant Effect of Pressure on Diver's Plasma. *J Proteome Res.* 2010; 9: 4131-4137.
53. Ferrarini A, Ruperez FJ, Erazo M, Martinez MP, Villar-Alvarez F, Peces-Barba G, et al. Fingerprinting-based metabolomic approach with LC-MS to sleep apnea and hypopnea syndrome: A pilot study. *Electrophoresis.* 2013; 34: 2873-2881.
54. Armirotti A, Basit A, Realini N, Caltagirone C, Bossu P, Spalletta G, et al. Sample preparation and orthogonal chromatography for broad polarity range plasma metabolomics: Application to human subjects with neurodegenerative dementia. *Anal Biochem.* 2014; 455: 48-54.
55. Postle AD. Phospholipid profiling. In: Griffiths WJ, editor. *Metabolomics, metabonomics and metabolite profiling.* 2007; p. 116-133.
56. Ciborowski M, Teul J, Martin-Ventura JL, Egido J, Barbas C. Metabolomics with LC-QTOF-MS permits the prediction of disease stage in aortic abdominal aneurysm based on plasma metabolic fingerprint. *PLoS One.* 2012; 7: e31982.

57. Rochat B. Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data. *J Am Soc Mass Spectrom.* 2017; 28: 709-723.
58. Koch A, Pfeilschifter J, Huwiler A. Sphingosine 1-Phosphate in Renal Diseases. *Cell Physiol Biochem.* 2013; 31: 745-760.
59. Bartels K, Grenz A, Eltzhig HK. Sphingosine-1-phosphate receptor signaling during acute kidney injury: the tissue is the issue. *Kidney Int.* 2014; 85: 733-735.
60. Van der Hooft JJJ, Barrett MP, Burgess KEV, Ridder L. Enhanced acylcarnitine annotation in high-resolution mass spectrometry data: fragmentation analysis for the classification and annotation of acylcarnitines. *Front Bioeng Biotechnol.* 2015; 3: 26.
61. Sirolli V, Rossi C, Di CA, Felaco P, Amoroso L, Zucchelli M, et al. Toward personalized hemodialysis by low molecular weight amino-containing compounds: future perspective of patient metabolic fingerprint. *Blood Transfus.* 2012; 10 Suppl 2: s78-88.
62. Mir S, Kantar M, Yalaz M, Keskinoglu A, Coker I, Huseyinov A. Effect of hemodialysis on carnitine levels in children with chronic renal failure. *Pediatr Int.* 2002; 44: 70-73.
63. Calvani M, Benatti P, Mancinelli A, D'Iddio S, Giordano V, Koverech A, et al. Carnitine replacement in end-stage renal disease and hemodialysis. *Ann N Y Acad Sci.* 2004; 1033: 52-66.
64. Walsh BH, Broadhurst DI, Mandal R, Wishart DS, Boylan GB, Kenny LC, et al. The metabolomic profile of umbilical cord blood in neonatal hypoxic ischaemic encephalopathy. *PLoS One.* 2012; 7: 1-12.
65. Maceyka M, Harikumar KB, Milstien S, Spiegel S. Sphingosine-1-phosphate signaling and its role in disease. *Trends Cell Biol.* 2012; 22: 50-60.
66. Awad AS, Rouse MD, Khutsishvili K, Huang L, Bolton WK, Lynch KR, et al. Chronic sphingosine 1-phosphate 1 receptor activation attenuates early-stage diabetic nephropathy independent of lymphocytes. *Kidney Int.* 2011; 79: 1090-1098.
67. Lee SB, Kanasaki K, Kalluri R. Circulating TGF- β 1 as a reliable biomarker for chronic kidney disease progression in the African-American population. *Kidney Int.* 2009; 76: 10-12.
68. Song JH, Kim M, Park SW, Chen SWC, Pitson SM, Lee HT. Isoflurane via TGF- β 1 release increases caveolae formation and organizes sphingosine kinase signaling in renal proximal tubules. *Am J Physiol.* 2010; 298: F1041-F1050.
69. Sell DR, Strauch CM, Shen W, Monnier VM. 2-Amino adipic acid is a marker of protein carbonyl oxidation in the aging human skin: effects of diabetes, renal failure and sepsis. *Biochem J.* 2007; 404: 269-277.
70. Benito S, Sanchez A, Unceta N, Andrade F, Aldamiz-Echevarria L, Goicolea MA, et al. LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease. *Anal Bioanal Chem.* 2016; 408: 747-760.

71. Andrade F, Rodriguez-Soriano J, Prieto JA, Aguirre M, Ariceta G, Lage S, et al. Methylation cycle, arginine-creatine pathway and asymmetric dimethylarginine in pediatric renal transplant. *Nephrol Dial Transpl.* 2011; 26: 328-336.
72. Fadel FI, Elshamaa MF, Essam RG, Elghoroury EA, El-Saeed GSM, El-Toukhy SE, et al. Some amino acids levels: glutamine, glutamate, and homocysteine, in plasma of children with chronic kidney disease. *Int J Biomed Sci.* 2014; 10: 36-42.
73. Sakoh T, Nakayama M, Tanaka S, Yoshitomi R, Ura Y, Nishimoto H, et al. Association of serum total bilirubin with renal outcome in Japanese patients with stages 3-5 chronic kidney disease. *Metab, Clin Exp.* 2015; 64: 1096-1102.
74. Tanaka M, Fukui M, Okada H, Senmaru T, Asano M, Akabame S, et al. Low serum bilirubin concentration is a predictor of chronic kidney disease. *Atherosclerosis.* 2014; 234: 421-425.
75. Boon AC, Lam AK, Gopalan V, Benzie IF, Briskey D, Coombes JS, et al. Endogenously elevated bilirubin modulates kidney function and protects from circulating oxidative stress in a rat model of adenine-induced kidney failure. *Sci Rep.* 2015; 5: 15482.



CHAPTER VI.

LC-QQQ-MS method for potential biomarkers

- 6.1. Introduction**
- 6.2. Study's goal**
- 6.3. Materials and equipment**
 - 6.3.1. Standards and reagents**
 - 6.3.2. Instrumentation**
 - 6.3.3. Preparation of standard solutions**
 - 6.3.4. Subjects and sampling**
- 6.4. LC-QQQ-MS method development**
 - 6.4.1. Selection of the chromatographic variables**
 - 6.4.2. Selection of the mass spectrometry detection variables**
 - 6.4.3. Optimization of sample treatment conditions**
- 6.5. Analytical evaluation of the method**
 - 6.5.1. Calibration of the method**
 - 6.5.2. Accuracy and precision**
 - 6.5.3. Stability**
- 6.6. Sample analysis**
- 6.7. Data analysis**
 - 6.7.1. Descriptive statistics**
 - 6.7.2. Univariate analysis**
 - 6.7.3. Multivariate analysis**
- 6.8. Conclusions**
- 6.9. Bibliography**

6.1. INTRODUCTION

There is an increasing interest in the development of simple methodologies that enable routine analysis for the diagnosis of different diseases in the clinical practice by means of a variety of endogenous biomarkers¹.

Regarding CKD and control paediatric differentiation, the untargeted metabolomics approach carried out in Chapter 5 showed increased ***n*-butyrylcarnitine**, ***cis*-4-decenoylcarnitine** and **sphingosine-1-phosphate** plasma levels, and decreased **bilirubin** plasma levels, which could be potential biomarkers.

However, scientists and clinicians concluded that untargeted metabolomics approaches are better suited for a discovery phase and should not be treated as a diagnostic tool in routine analysis to differentiate between healthy and disease conditions of different populations. Indeed, the end result of untargeted metabolomics studies are statistically significant up- or down-regulated compounds that can give an insight of the progress or pathogenesis of specific conditions. However, to find out which of the identified significant metabolites might act as biomarkers, there is a need for analytical method validation². Analytical method validation aims at confirming that an analytical procedure used for a specific test is suitable for its intended use. For that purpose, analytical requirements are defined to assure that the method under consideration has performance capabilities consistent with the requirement of the application.

The reliability of the results obtained by means of multi-instrument inter-laboratory untargeted metabolomics studies has been recently assessed³. According to this research work, there is a high level of analytical convergence regarding the use of multiple instruments in different locations, but there is still a need to validate the biological outcomes arising from untargeted metabolomics analyses. The validation of the biological results obtained from untargeted metabolomics studies needs to be carried out developing and validating targeted specific analytical methods for the specific significant metabolites obtained². Additionally, the use of multisite clinical trials is becoming increasingly important in the validation process, due to the fact that some factors, such as diet, environment, race or socioeconomic factors may have a geographic origin⁴.

Despite the fact that multiple parameters were controlled during untargeted metabolomics study carried out aimed at paediatric CKD to assure the stability of the system, there is a

need for validating these metabolites as potential biomarkers by means of a quantitative analytical method.

Apart from the metabolites obtained from untargeted metabolomics assay, **citrulline**, **S-dimethylarginine**, **S-adenosylmethionine** and **creatinine** were found to be significant in the targeted metabolomics study performed to find up- or down-regulated metabolites in arginine-creatine, urea cycle and arginine methylation metabolic pathways (Chapter 4)^{5,6}.

Combination of the potential biomarkers in one common routine analysis method could be of interest to assess the metabolite pattern in both control and CKD paediatrics and could be of application in screening routine analyses carried out by general practitioners. Indeed, it is estimated that the majority of the population visits their general practitioner at least once within a 3-year period and can be subjected to screening⁷.

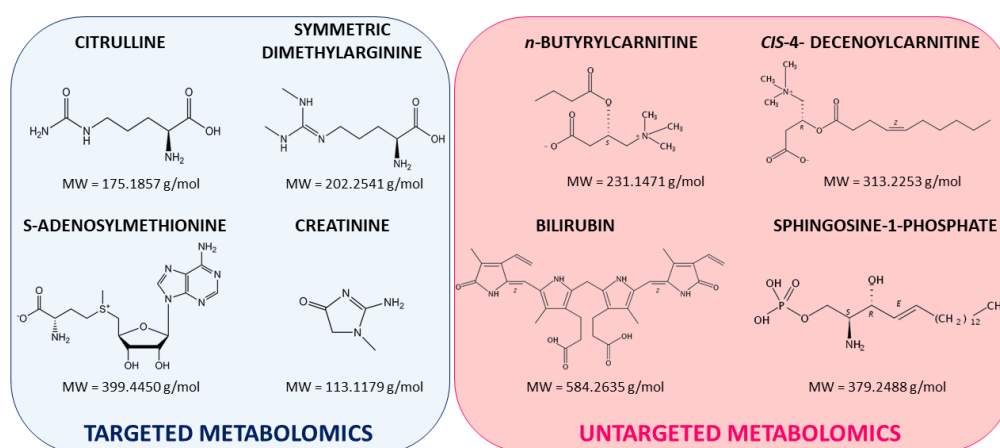


Fig. 6.1. Potential biomarkers according to targeted and untargeted metabolomics studies carried out and showed in Chapters 4 and 5.

6.2. STUDY'S GOAL

The main objective of this study deals with developing and validating a routine quantitative analytical method, which gathers the analytes found to be significant in both targeted metabolomics study described in *Chapter 4* and untargeted metabolomics study from *Chapter 5*. This new analytical method will accomplish different goals:

- ✓ Validation of the biological results obtained from the untargeted metabolomics study from Chapter 5.
- ✓ Obtaining a new LC-MS method that collects significant metabolites for CKD and control population discrimination from the targeted and untargeted metabolomics studies aimed at quantification.
- ✓ Being simple for its potential application in clinical practice.

To accomplish these purposes, a routine analysis method needs to be carried out using LC-QQQ equipment. This method will aim at the quantification of the following analytes: citrulline (CIT), symmetric dimethylarginine (SDMA), S-adenosylmethionine (SAM), creatinine (CNN), *n*-butyrylcarnitine (nC4), *cis*-4-decenoylcarnitine (CIS4DEC), sphingosine-1-phosphate (S1P) and bilirubin (BIL). Both extraction method and analysis method optimization will be required to obtain the highest sensitivity, followed by the validation of the method, aimed at assuring the quality of data obtained.

6.3. MATERIALS AND EQUIPMENT

6.3.1. STANDARDS AND REAGENTS

Methanol and acetonitrile used in the plasma extraction study were acquired from Scharlab (Barcelona, Spain). Besides, LC-MS grade formic acid used for mobile phase was obtained from Fisher Scientific (Ghent, Belgium), ammonium formate from Fluka Analytical, Sigma-Aldrich (Steinheim, Germany) and perfluoroheptanoic acid 96 % from Across Organics (New Jersey, USA). Ultra-high purity water achieved from tap water pretreated by Elix reverse osmosis and a Milli-Q system from Millipore (Bedford, MA, USA) was used. Moreover, mobile phases were filtered through 0.1 μm filters from Millipore Omnipore (Watford, Ireland) prior to use.

The analytical standards used were acquired from different manufacturers. *N*_G,*N*_G'-dimethyl-L-arginine di(p-hydroxyazobenzene-p'-sulfonate) (SDMA), S-adenosyl-L-methionine (SAM), *n*-butyryl-L-carnitine (nC4), *iso*-butyryl-L-carnitine (iC4) and citrulline (CIT) were acquired from Sigma-Aldrich (Steinheim, Alemania). In addition, creatinine (CNN) was obtained from Alfa Aesar (Karlsruhe, Germany) manufacturer, D-erythro-sphingosine-1-phosphate (S1P) was purchased from Larodan AB (Limhamn, Sweden), *cis*-4-decenoylcarnitine synthesized by Lumila Research group at The Autonomous University of Madrid (Madrid, Spain) and bilirubin was obtained from TCI (Tokyo, Japan).

Regarding isotopically labeled compounds, N_G, N_G' -dimethyl-L-arginine- d_6 (SDMA- d_6) and creatinine- d_3 (CNN- d_3) were both supplied by Toronto Research Chemicals, TRC-Canada (North York, Canada).

In addition, blood sample collection was carried out using Becton Dickinson EDTA tubes (Plymouth, UK).

6.3.2. INSTRUMENTATION

The analysis method gathering the eight metabolites of interest was accomplished injecting 2 μ L of sample in a 1200 series HPLC system coupled to a 6410 series triple quadrupole mass spectrometer (LC-QQQ) from Agilent Technologies (Santa Clara, CA, USA), equipped with Agilent Jet Stream electrospray source. Chromatographic separation was carried out using the reversed phase column Zorbax Eclipse Plus C18 (3.0 x 50 mm, 1.8 μ m) and C8 guard column (2.1 x 12 mm, 5 μ m), both from Agilent Technologies, which were kept at 40 °C during the analysis. In addition, another two columns were also assayed, but were discarded because of their bad performance: Kinetex C18 100 Å (2.1 x 30 mm, 1.7 μ m) and Zorbax Extend-C18 Rapid resolution HT (2.1 x 50 mm, 1.8 μ m).

Data acquisition was carried out using Agilent MassHunter Workstation Data Acquisition version B.08.00 software and afterwards, raw data was processed using MassHunter Qualitative Analysis version B.07.00 (both from Agilent Technologies). Finally, data analysis was performed using Matlab R2015a (Mathworks) and SPSS Statistics 23 (IBM).

Samples were centrifuged with a 5415 R Eppendorf (Madrid, Spain) centrifuge, and supernatant evaporation was carried out in a Techne, Dri-Block® DB-3D (Staffordshire, UK) evaporator system.



Fig. 6.2. LC-QQQ instrument used in this study.

6.3.3. PREPARATION OF STANDARD SOLUTIONS

First of all, stock standard solutions containing around 1000 mg/L were prepared and from these solutions individual standard solutions of 1 mg/L, 2 mg/L and 5 mg/L were made for precursor ion identification. Stock solution of CNN was prepared in water, whereas CIT and SAM were dissolved in 0.01 N of HCl, to contribute to their stability. Besides, SDMA, CIS4DEC, nC4, nC4 and S1P were prepared in methanol and BIL in acetonitrile. Concerning internal standards, creatinine-d₃ and SDMA-d₆ were dissolved in methanol. These solutions were stored frozen at -40°C.

Intermediate mixed solutions containing BIL and SAM were prepared in acetonitrile and were replaced twice a week, whereas the rest of the analytes were made weekly in methanol. Working solutions were obtained from intermediate mixed solutions for analytical methodology and sample treatment optimization. Due to the observed degradation of working solutions at room temperature, there was a need for replacing them daily with frozen intermediate solution.

Calibration standards were prepared spiking pooled plasma made of control samples, which had been previously quantified by standard addition and contained: 5.9 µg/mL CIT, 9.4 µg/mL CNN, 145 ng/mL SDMA, 19 ng/mL nC4, 40 ng/mL SAM, 75 ng/mL CIS4DEC, 40 ng/mL S1P and 7.5 µg/mL BIL. After being doped, both calibration standards and real samples underwent the same sample treatment.

Quality control samples were prepared using a pooled plasma previously quantified by means of standard addition. This pooled plasma contained 5.3 µg/mL CIT, 7.1 µg/mL CNN, 132 ng/mL SDMA, 18 ng/mL nC4, 61 ng/mL SAM, 71 ng/mL CIS4DEC, 210 ng/mL S1P and 8.5 µg/mL BIL. Low, medium and high QC samples were obtained spiking this pool. High QC consisted in the addition of 9 µg/mL of CIT, 22.5 µg/mL of CNN, 90 ng/mL of SDMA, 22.5 ng/mL nC4, 600 ng/mL of SAM, 225 ng/mL CIS4DEC, 4.5 µg/mL of S1P and 11.25 µg/mL of BIL. Regarding low and medium QC, same pooled plasma was used but spiked with the 20 % and the 50 % of the concentration added for each analyte in high QC level. These QC samples underwent the same sample treatment as calibration standards and samples, and were analyzed during the batch to check the correct performance of the analytical method and the accuracy of the quantifications carried out.

6.3.4. SUBJECTS AND SAMPLING

The routine analytical method developed was applied to plasma samples from thirty-two patients suffering from CKD, aged 3-18 years, and thirteen control patients, aged 6-16 years. Samples from CKD patients were the same previously used in Chapter 4 and 5, whereas, control samples differed in some cases, but followed the same inclusion criteria: healthy children who had a minor surgery in Cruces University Hospital. Paediatrics suffering from CKD were followed up in Cruces University Hospital (Bilbao, Spain) and had been classified into different degrees of the disease, as showed in Table 6.1.

Table 6.1. Characteristics of the plasma samples quantified with this routine method.

CKD DEGREE	Characteristics of the population			Number of patients
	SEX (M/F)	AGE (2-12 y/13-18 y)	TREATMENT (No RRT/ Dialyzed/ Transplanted)	
Control	11/3	11/2	13/0/0	13
CKD2	8/6	10/4	9/0/5	14
CKD3	5/1	2/4	4/0/2	6
CKD4	2/4	3/3	5/0/1	6
CKD5	2/4	5/1	1/5/0	6

Blood samples were collected in the morning after an overnight fasting. Samples were immediately cooled in an ice-water bath and centrifuged at 1000 g for 5 min at 4°C. Finally, they were stored at -80°C until sample treatment and analysis.

Ethics Committee of Clinic Research of Cruces Hospital approved the study protocol and patients' parents gave written informed consent.

6.4. LC-QQQ-MS METHOD DEVELOPMENT

6.4.1. SELECTION OF THE CHROMATOGRAPHIC VARIABLES

First attempts in analytical method optimization were carried out in **MS2 Scan mode**. Due to the nature of all of these analytes ESI interface was used and first attempts were made using 0.1 % formic acid and 0.5 mM PFHA aqueous phase (A) and acetonitrile organic phase (B).

The eight analytes under study as well as the two internal standards used (CNN-d₃ and SDMA-d₆) were analysed and characterized, one by one, obtaining their retention times

and their characteristic m/z . For all the metabolites, the protonated form of each of them, $[M+H]^+$, was obtained, as showed in Table 6.2.

Table 6.2. Molecular formulas, molecular masses, ionization and experimental m/z ratios in MS for analytes of interest.

Analyte	Molecular formula	MW	Ionization	m/z
CIT	$C_6H_{13}O_3N_3$	175.1	$[M+H]^+$	176.1
CNN	$C_4H_7ON_3$	113.1	$[M+H]^+$	114.1
CNN-d ₃	$C_4H_4D_3ON_3$	116.1	$[M+H]^+$	117.1
SDMA	$C_8H_{18}O_2N_4$	202.1	$[M+H]^+$	203.2
SDMA-d ₆	$C_8H_{12}D_6O_2N_4$	208.1	$[M+H]^+$	209.2
n-C4	$C_{11}H_{21}NO_4$	231.2	$[M+H]^+$	232.2
SAM	$C_{15}H_{22}O_5N_6S$	398.1	$[M+H]^+$	399.1
CIS4DEC	$C_{17}H_{31}NO_4$	313.2	$[M+H]^+$	314.2
S1P	$C_{18}H_{38}NO_5P$	379.3	$[M+H]^+$	380.3
BIL	$C_{33}H_{36}N_4O_6$	584.3	$[M+H]^+$	585.3

As showed in Chapter 5, in addition to *n*-butyrylcarnitine (nC4), its structural isomer *i*-butyrylcarnitine (iC4) is also expected in plasma. Due to the fact that both ionize equally in MS and MS/MS mode, for method optimization both nC4 and iC4 will be monitored, to assure that these metabolites are separated in the chromatogram. However, once developed the method, only nC4 will be included on it, because iC4 did not showed to be significant according to the study carried out in Chapter 5.

6.4.1.1. Column

S1P sphingolipid is a particular analyte, which contains a long hydrocarbon chain in addition to hydroxyl, amino and phosphate groups, and a pKa value of 1.76⁸. Preliminary studies carried out using 0.1 % formic acid in the aqueous phase (A) and acetonitrile in the organic phase (B) showed that this compound may present problems to elute from column. Aimed at verifying that the problem did not come from the analytical standard or from poor ionization, direct injection was experimented using 0.1 % formic acid (A) and acetonitrile (B), in isocratic mode, with different proportions of solvents: 80 % B, 50 % B and 20 % B. In all the different isocratic methods performed by direct infusion of stock solutions, S1P provided a signal. Therefore, problems with the analytical standard solution or with poor ionization were discarded.

Different C18 chromatographic columns were assayed in order to find one that achieves S1P elution in addition to obtaining a good relation between separation of the analytical compounds and peak resolution: Kinetex C18 100 Å (2.1 x 30 mm, 1.7 µm), Zorbax Extend-C18 Rapid resolution HT (2.1 x 50 mm, 1.8 µm) and Zorbax Eclipse Plus C18 Rapid resolution (3 x 50 mm, 1.8 µm).

First of all, multiple gradients and mobile phases were experimented with **Kinetex C18** column. However, none of the conditions assayed enabled S1P elution.

Thereafter, **Zorbax Extend-C18** column, which aims at separating basic compounds above their pKa in their free base form, was experimented. In this case, mobile phase flow was reduced from 0.3 mL/min to 0.25 mL/min, as it is the maximum recommended flow for that column. Nevertheless, again no signal was observed for S1P.

Finally, **Zorbax Eclipse Plus C18**, which is intended for the separation of acidic, basic and other highly polar compounds by reverse-phase liquid chromatography was assayed with the same mobile phase with a flow rate of 0.3 mL/min. In this case, signal was obtained for all the analytes and peaks showed an acceptable resolution.

6.4.1.2. Composition of the mobile phase

Once selected Zorbax Eclipse Plus C18 column, composition of the mobile phase was optimized. With this column at 30 °C, different gradients using 0.1 % formic acid (A) and acetonitrile (B) were assayed to check whether a good separation of the analytes could be achieved.

Even when the organic phase proportion was minimal (2 % B) and regardless of the gradients assayed, CIT, CNN, SDMA, iC4, nC4 and SAM elute together at the start of the chromatogram (around 0.7 minutes). Regarding that in Chapter 4, PFHA ion-pairing reagent was added in the aqueous phase providing better retention to charged polar compounds, the use of this reagent was considered (see Fig. 6.3). Again a concentration of 0.5 mM was used matching with previous literature⁹⁻¹¹ and experience from Chapter 4.

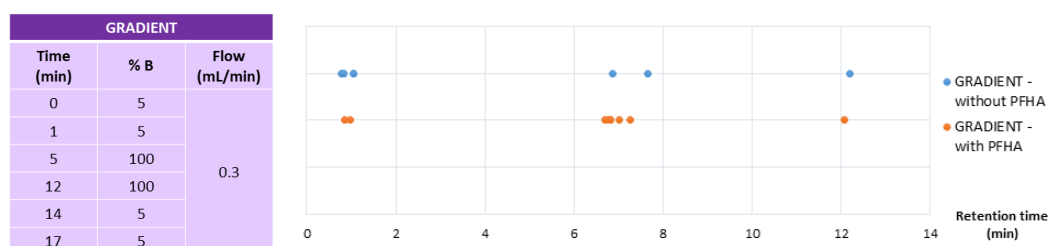


Fig. 6.3. Differences in analyte separation using gradient on the left with a mobile phase containing and not containing PFHA ion pairing reagent.

The addition of PFHA to the aqueous phase provided a significant improvement in the separation of the analytes, as it can be inferred from Fig. 6.3. Therefore, using PFHA to improve the retention was decided.

Provided that this method gathers analytes with different nature, the effect of using different mobile phase modifiers was studied. The addition of either 5 mM ammonium formate or 0.1 % formic acid to 0.5 mM PFHA containing aqueous phase was experimented, obtaining better results with formic acid modifier. Besides, 0.5 mM of PFHA and 0.1 % formic acid were added to the organic phase, due to the fact that this mobile phase provided better results in terms of between sample retention time repetitivity and ionization of S1P and SAM. Thus, the selected mobile phase contained 0.1 % formic acid and 0.5 mM PFHA (A) and 0.1 % formic acid and 0.5 mM PFHA in acetonitrile (B).

Next step consisted in improving the separation of the analytes. Aimed at obtaining a better separation of the metabolites, initial percentages of organic phase to 2 %, 5 %, 10 %, 15 % and 20 % B were assayed. Differences found after testing these gradients are showed in Fig. 6.4.

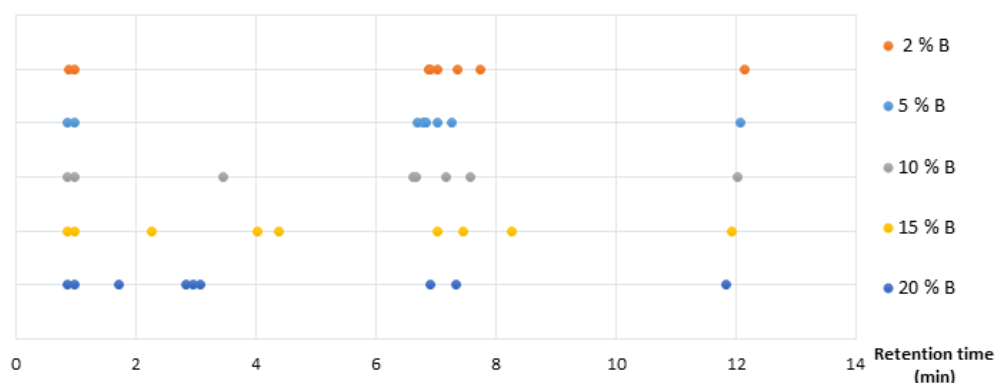


Fig. 6.4. Comparison of starting percentages of organic phase for the same gradient.

Gradient starting from 15 % B seems to provide the best separation for all the metabolites. However, as it has been previously stated, it is necessary to assess the separation between iC4 and nC4 as well. Separation of these analytes with previous gradients was compared and is showed in Fig. 6.5. Taking into account the separation of the isomers and their peak shape, the best separation was achieved using gradients starting from 20 %B.

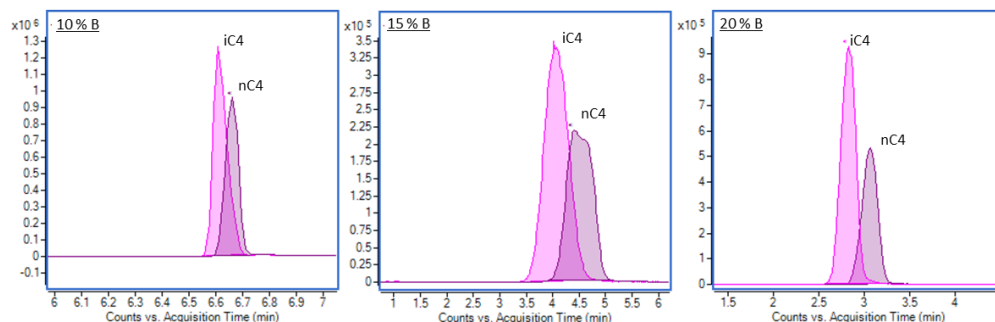


Fig. 6.5. Scan EIC ($232.2 m/z$) chromatograms of iC4 and nC4 superimposed showing the effect of starting with different organic percentages on butyrylcarnitine isomer separation.

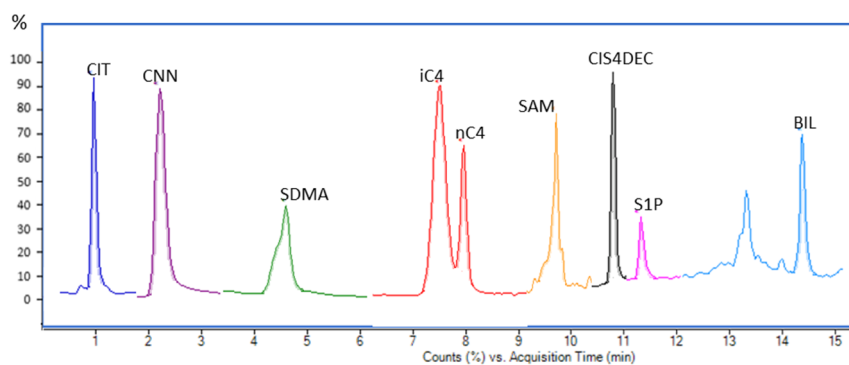
At this point column temperature was optimized experimenting with 30, 35 and 40 °C. At a temperature of 40 °C, better signals were obtained for SDMA, SAM, nC4, iC4, S1P and BIL, whereas similar peaks were obtained in all the temperature range for CIT, CNN and CIS4DEC. Moreover, in case of BIL, peak was more symmetric with 40 °C and less tailing effect was observed. For all these reasons, a column temperature of 40 °C was selected.

Finally, with the aim of obtaining a better separation, the gradual increase of the organic phase was slowed down and at the time iC4 and nC4 eluted, flow was suddenly decreased. Definitive chromatographic conditions, including gradient program are showed in Table 6.3.

Table 6.3. Definitive chromatographic conditions selected for this LC-QQQ-MS method.

Chromatographic conditions			
Column	Zorbax Eclipse Plus C18 Rapid resolution (3 x 50 mm, 1.8 μ m) C8 guard column (2.1x12 mm, 5 μ m)		
Column temperature	40 \pm 0.8 $^{\circ}$ C		
Mobile phase	A: 0.1 % formic acid and 0.5 mM PFHA B: 0.1 % formic acid and 0.5 mM PFHA in acetonitrile		
Gradient program	Time (min)	% B	Flow (mL/min)
	0	20	0.4
	1	20	-
	4	30	-
	6	30	0.4
	6.1	30	0.15
	7	30	0.15
	7.1	30	0.4
	7.5	100	-
	15	100	-
16.5	20	-	
20	20	0.4	
Injection volume	2 μ L, needle wash with isopropanol:water (50:50, v/v)		

Once the definitive composition of the mobile phase was set, analytical standards were analyzed with the m/z values selected for each analyte (see Table 6.2) in **MS2 selected ion monitoring (SIM) mode** for further sensitivity, and the chromatogram showed in Fig. 6.6 was obtained.

**Fig. 6.6.** Normalized SIM chromatogram obtained from the analysis of 2 μ g/mL of each of the analytes of interest.

6.4.2. SELECTION OF THE MASS SPECTROMETRY DETECTION VARIABLES

Once the separation conditions of the analytes had been defined, different variables from the mass spectrometer were optimized to increase the sensitivity of the analytes of interest.

6.4.2.1. MS spectra

Taking into account that all of these analytes ionized in electrospray (ESI) according to previous experiments (see Chapter 4 and Chapter 5), ESI positive mode was used for the analysis of these metabolites.

First of all, different mass spectrometry conditions were assayed and optimized for the analytes of interest: capillary voltages (V_{cap}) of 3000 V, 3500 V and 4000 V; fragmentor voltages (Frg) of 75 V, 100 V, 125 V and 150 V; drying gas temperatures of 200 °C, 250 °C, 300 °C and 325 °C; and drying gas flows of 9 L/min, 10 L/min, 11 L/min, 12 L/min. A compromise was reached to select the conditions that improve those analytes with less signal, affecting as less as possible to the rest of the metabolites. Optimized mass spectrometry conditions are summarized in Table 6.4.

Table 6.4. Optimized MS spectrometry conditions for LC-QQQ-MS method.

Mass spectrometry conditions	
Interface	ESI
Fragmentor voltage	100 V
Dry gas	250 °C, 10 L/min
Nebulizer pressure	40 psig
V_{cap}	3500 V
Mass range	SCAN: 100-1000 m/z

6.4.2.2. MS/MS spectra

Once selected MS conditions, **product ion scan** analysis was carried out to obtain characteristic transitions to be used for quantification and qualification. Different collision energies were experimented (10 V, 20 V, 30 V and 40 V), being selected those with the most significant spectra. Fig. 6.7 shows the product ion scans obtained from all the analytes, including the isotopically labelled compounds.

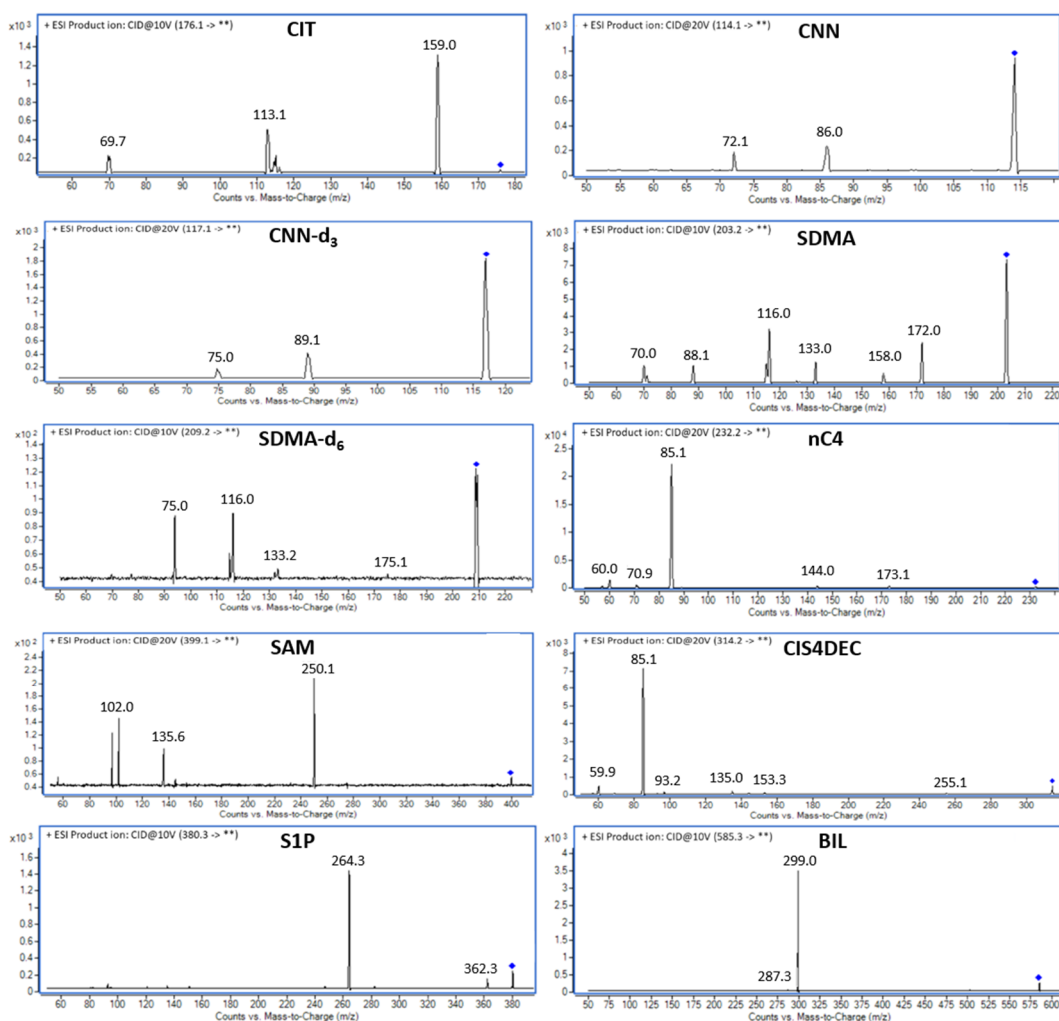


Fig. 6.7. Product ion spectra of the targeted analytes.

As it can be observed from this figure, for CNN, CNN-d₃, SDMA and SDMA-d₆ precursor ion was still abundant with the collision energies selected. The reason why a collision energy of 20V is used for CNN and CNN-d₃ instead of a higher voltage, regarding that precursor ion was still abundant, is that collision energies of 25 V and 30 V were assayed and the fragmentation was too high to obtain any significant product ion. On the other hand, SDMA has a structural isomer, ADMA, which elutes at the same retention time and has the same MS spectra. However, according to literature¹²⁻¹⁴ and to our previous experience (see Chapter 4) their MS/MS fragmentation at low collision energies provides unique and specific ion products. After assaying different energies, 10 V provided a unique product ion, 172.1 *m/z*, for SDMA. Accordingly, SDMA-d₆ was analyzed with 10 V as well, with the aim of obtaining an analogous behaviour.

EU's decision 2002/657/CE states the minimum requirements for the collection of identification points for different analytical methodologies¹⁵. For LC-MS/MS methodologies a minimum of 4 identification points are needed. This criterion is fulfilled using 1 precursor ion and 2 product ions at a defined retention time. Following this statement, two different transitions were selected, one for quantification (Q) and the other one for qualification (q) as showed in Table 6.5.

According to the selected quantification and qualification transitions **dynamic multiple reaction monitoring (dMRM)** mode was set.

Table 6.5. MRM transitions selected for quantification and qualification.

Dynamic MRM	RT (min)	Transition	EC (V)
CIT	1.09	176.1 > 159 (Q) 176.1 > 113.1 (q)	10
CNN	1.76	114.1 > 86 (Q) 114.1 > 72.1 (q)	20
CNN-d ₃	1.76	117.1 > 89.1 (Q) 117.1 > 75 (q)	20
SDMA	3.50	203.2 > 172.1 (Q) 203.2 > 158 (q)	10
SDMA-d ₆	3.50	209.2 > 175.1 (Q) 209.2 > 116.1 (q)	10
n-C4	8.35	232.2 > 85.1 (Q) 232.2 > 173.1 (q)	20
SAM	9.06	399.1 > 250.0 (Q) 399.1 > 135.9 (q)	20
CIS4DEC	10.98	314.2 > 85.1 (Q) 314.2 > 255.1 (q)	20
S1P	11.38	380.3 > 264.3 (Q) 380.3 > 362.3 (q)	10
BIL	14.62	585.3 > 299 (Q) 585.3 > 287.3 (q)	10

The analysis of a stock solution containing the metabolites of interest using MRM transitions showed in Table 6.5 lead to the following chromatograms:

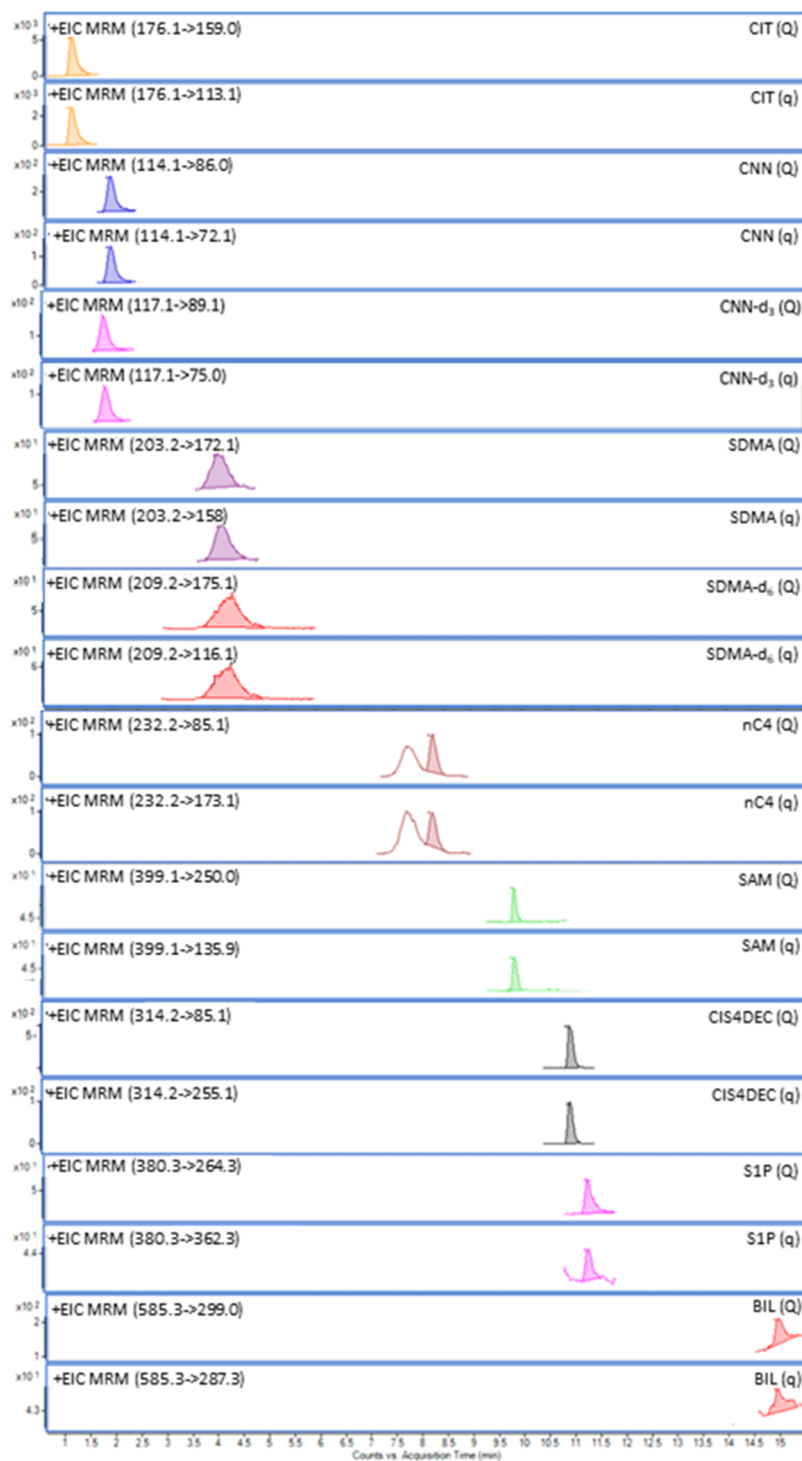


Fig. 6.8. MRM transitions of stock solution prepared in acetonitrile:water (50:50, v/v) containing 12 $\mu\text{g/mL}$ CIT, 30 $\mu\text{g/mL}$ CNN, 15 $\mu\text{g/mL}$ CNN- d_3 , 120 ng/mL SDMA, 100 ng/mL SDMA- d_6 , 30 ng/mL iC4, 30 ng/mL nC4, 200 ng/mL SAM, 300 ng/mL CIS4DEC, 1.5 $\mu\text{g/mL}$ S1P and 15 $\mu\text{g/mL}$ BIL.

6.4.3. OPTIMIZATION OF SAMPLE TREATMENT CONDITIONS

The optimization of sample treatment procedures enables both extracting targeted analytes and decreasing any matrix effect in order to obtain low limits of quantification.

6.4.3.1. Addition of isotopically labelled compounds

The use of isotopically labelled compounds is useful to correct for differences suffered by samples during sample treatment. In this case, two different isotopically labeled compounds were used. Creatinine-d₃ (CNN-d₃) was used to correct for the concentrations of CIT and CNN, whereas SDMA-d₆ corrected the signals of SDMA, nC4, SAM, CIS4DEC, S1P and BIL. 10 µL of a mix solution containing 100 µg/mL of CNN-d₃ and 1 µg/mL of SDMA-d₆ was added to 50 µL of plasma, to reach concentrations of 20 µg/mL of CNN-d₃ and 200 ng/mL of SDMA-d₆.

6.4.3.2. Protein precipitation reagent

Protein precipitation (PPT) is the simplest approach and performed well for the analytes of interest as seen in Chapters 4 and 5. For that reason, PPT was carried out. Regarding that in the untargeted metabolomics method developed methanol:ethanol (50:50, v/v) was used for PPT, this solvent was first experimented. However, even if this solvent enabled the extraction of all the analytes of interest, and that BIL was also extracted with this solvent mixture, when increasing BIL concentration to build calibration curves in plasma, no linearity was found. According to literature BIL is soluble in apolar solvents, such as chloroform (10 g/L), practically insoluble in water and soluble in aqueous solutions at high pH (300 g/L)^{8,16}. However, it has to be noted that alkaline solutions rapidly oxidize BIL¹⁷. Therefore, the hypothesis that BIL extraction would be dependent on its solubility in the precipitation reagent was proposed to explain that lack of linearity. Consequently, the use of other solvents with different polarities was studied.

The following solvents were experimented for PPT as precipitation reagents by adding to 50 µL of pooled plasma 150 µL of the following frozen solvents and were compared with results from methanol:ethanol (50:50, v/v) extraction: acetonitrile, 1 % formic acid in acetonitrile, methanol and 1 % formic acid in methanol. Afterwards, all the extracts were evaporated and reconstituted in methanol. Regarding that recoveries for each analyte cannot be calculated as there is not any analyte-free matrix available, the relation between areas of the analytes of interest normalized with the signal of internal standards for

different precipitation reagents on average were compared. Results from this study have been summarized in Fig. 6.9.

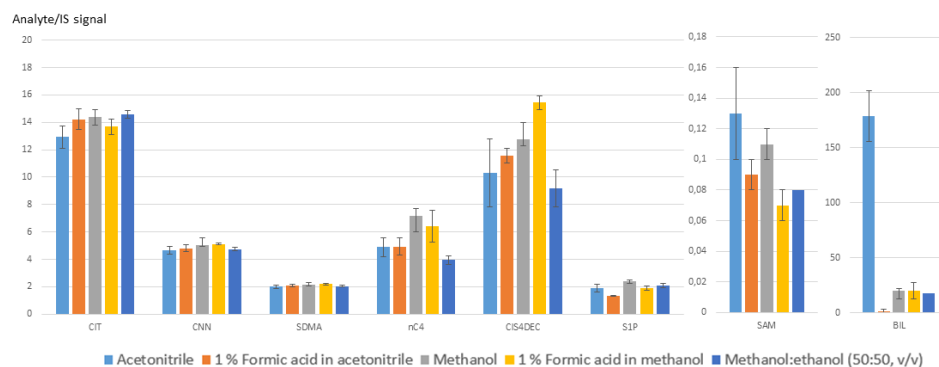


Fig 6.9. Analyte/internal standard (IS) ratios and standard deviations obtained from plasma extraction with acetonitrile, 1 % formic acid in acetonitrile, methanol, 1 % formic acid in methanol and methanol:ethanol (50:50, v/v).

In general, the extraction was similar for all the metabolites using any of the precipitation reagents. However, for BIL significant differences were found. Indeed, plasma precipitation and extraction of BIL with acetonitrile provided 10 times higher extraction in comparison with the use of methanol, 1% formic acid in methanol and methanol:ethanol (50:50, v/v), and 100 times higher extraction in comparison with 1 % formic acid in acetonitrile. Moreover, plasma extracts obtained from acetonitrile showed a yellowish colour, whereas the rest of the extracts were almost transparent, which may give us a clue about the recovery of the extraction, taking into account that bilirubin solutions in acetonitrile are yellow. The poor extraction using 1 % formic acid in acetonitrile could be related to the acidity of the solution, since evidences have been found of its slight solubility in extremely acid aqueous solutions⁸.

Regarding all these concerns, even though acetonitrile was discarded for PPT in the untargeted study in Chapter 5, because it increases phospholipid extraction, acetonitrile was selected for plasma protein precipitation in this LC-QQQ-MS method. Indeed, in this case a targeted methodology is being developed, which means that each analyte is individually monitored according to specific transitions, and a potential increase of total ion current chromatogram as a consequence of phospholipid accumulation would not affect the quantification in MRM.

Due to the fact that metabolites of interest are expected in different concentration ranges and regarding that the polarity of these analytes differs, sample evaporation and reconstitution of plasma extracts need to be optimized.

PPT was assayed using acetonitrile, evaporating the extracts in nitrogen stream and reconstituting them in 100 μ L of the following solvents: acetonitrile, 1 % formic acid in acetonitrile, methanol, initial conditions of the mobile phase, acetonitrile:acetic acid 0.5 M (75:25, v/v) and acetonitrile:acetic acid 0.1 M (75:25, v/v). In addition, the effect of avoiding any evaporation and reconstitution, by analysing sample extract obtained from acetonitrile PPT directly was also experimented.

When using either acetonitrile or 1 % formic acid in acetonitrile some metabolites, such as, CIT, CNN and SDMA, were not detected in control plasma. Therefore, the use of acetonitrile or 1 % formic acid in acetonitrile was discarded. The results for the rest of the conditions studied are showed in Fig. 6.10.

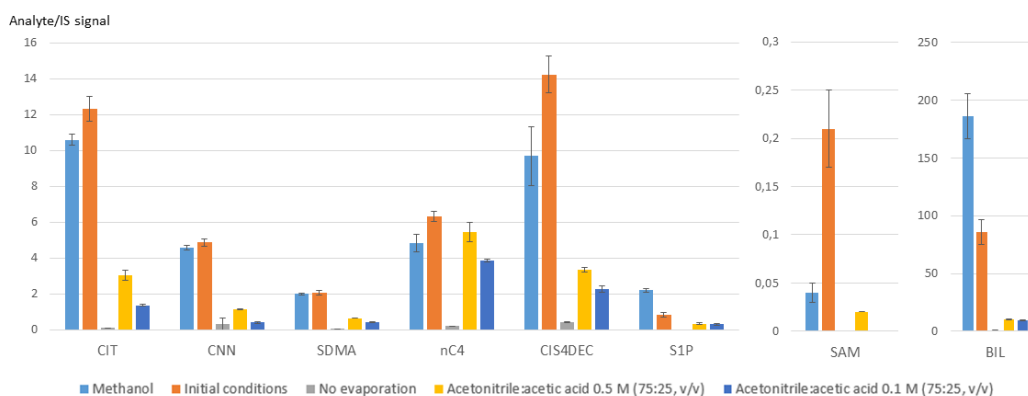


Fig. 6.10. Analyte/IS ratios and standard deviations obtained from plasma extraction with acetonitrile and reconstitution in methanol, initial mobile phase conditions, acetonitrile:acetic acid 0.1 M (75:25, v/v), acetonitrile:acetic acid 0.5 M (75:25, v/v) and without any evaporation and reconstitution process.

Best recoveries of the analytes were obtained using methanol and initial conditions of the mobile phase. However, as time goes by BIL and SAM signal decreased. In contrast, despite obtaining worse recoveries with acetonitrile:acetic acid 0.5 M, the stability of all the compounds was better. This would match with literature advising the use of acetic acid to acidify plasma for a better stability of SAM¹⁸. Regarding that the stability was improved and that the extracted amount of analyte was considered enough for quantification of these analyte in plasma, acetonitrile:acetic acid 0.5 M (75:25, v/v) was selected for reconstitution.

Taking into account that on the one hand, some analytes like CNN or BIL might be present in plasma above the linearity range and on the other hand, other analytes like SAM, nC4, SDMA and CIS4DEC were expected within ng/mL concentration range, it was considered necessary the optimization of reconstitution volume. Different reconstitution volumes of acetonitrile:acetic acid 0.5 M (75:25, v/v) were assayed: 50 μ L, 75 μ L, 100 μ L and 200 μ L. Finally, reaching a compromise, a volume of 100 μ L was selected.

After optimizing all the steps of sample preparation, plasma pretreatment method was set as follows:

1. Plasma was thawed at room temperature
2. 50 μ L of each plasma sample were placed in Eppendorf tubes
3. 10 μ L of internal standard mixture containing 100 μ g/mL of creatinine-d₃ and 1 μ g/mL of SDMA-d₆ were included
4. Protein precipitation was carried out by means of the addition of 150 μ L of frozen acetonitrile
5. Samples were vortexed and subsequently centrifuged at 13000 rpm for 10 min at 4°C
6. The supernatant obtained was evaporated in nitrogen stream and reconstituted in 100 μ L of solution containing 75 % acetonitrile and 25 % acetic acid 0.5 M

6.5. ANALYTICAL EVALUATION OF THE METHOD

6.5.1. CALIBRATION OF THE METHOD

Calibration curve is ideally obtained spiking analyte-free matrix with known quantities of the analytes of interest. However, as previously explained in Chapter 4, there is not any commercially available plasma matrix free of all of these analytes. Moreover, diverse surrogates were tested for some of the metabolites of interest and resulted into a different matrix effect. Therefore, regarding that the available plasma volume did not allow quantification of individual samples using standard addition method, calibration curves were prepared in pooled plasma previously quantified by standard addition and spiked with known concentrations of the analytes. It has to be noted that according to the results obtained from targeted and untargeted metabolomics methods developed in Chapters 4 and 5, for all the metabolites except for BIL increased concentrations were expected in CKD paediatrics. Therefore, calibration curves were built in pooled plasma containing only

control samples, since lower concentrations of these compounds were expected in these samples.

Regression equations of these 8 analytes are summarized in Table 6.6. These equations were obtained normalizing analyte signal with internal standards and facing this ratio with abundances from areas of MRM chromatograms. In this case, correlation coefficients above 0.9979 were achieved.

Table 6.6. Equations used for quantification of the metabolites in plasma and statistical analysis on the random distribution of residues.

	Calibration range ($\mu\text{g/mL}$)	Slope \pm error	Intercept \pm error	Correlation coefficient (r)	σ_{TOT}	\bar{r}	$\sigma(r)$	$ \bar{r} $
CIT	5.9-53.9	0.23 ± 0.02	1.34 ± 0.28	0.9990	0.050	0.000	0.039	0.030
CNN	9.4-129.4	0.04 ± 0.01	0.38 ± 0.07	0.9995	0.045	0.000	0.035	0.026
SDMA	0.145-0.625	4.12 ± 0.66	0.59 ± 0.12	0.9979	0.031	0.000	0.024	0.022
nC4	0.019-0.139	86.36 ± 17.82	1.64 ± 0.39	0.9997	0.051	0.000	0.040	0.036
SAM	0.040-0.540	0.42 ± 0.05	0.01 ± 0.00	0.9992	0.004	0.000	0.003	0.003
CIS4DEC	0.075-0.375	44.95 ± 8.03	3.38 ± 0.76	0.9992	0.077	0.000	0.060	0.048
S1P	0.040-1.5	0.84 ± 0.19	0.26 ± 0.09	0.9988	0.020	0.000	0.015	0.013
BIL	7.5-37.5	2.35 ± 0.16	17.84 ± 4.61	0.9989	0.764	0.000	0.540	0.522

The adequate adjustment of linear regression equations obtained to data needs to be verified by means of a statistical analysis on the random distribution of the residues. As previously explained in Chapter 4, to comply with the assumption that linear regression adjusts to data contained, standard deviation of the residues ($\sigma(r)$) needs to be equal to or less than global standard deviation (σ_{TOT}) and average residue ($|\bar{r}|$) equal to or lower than $\sigma(r)$. Besides, if data distribution is normal, the arithmetic mean of the residues (\bar{r}) is expected to be close to 0. As it can be observed from Table 6.6, all the assumptions explained above are fulfilled. Thus, it is concluded that the equations of linear regression are adequate to quantify plasma samples.

6.5.2. ACCURACY AND PRECISION

Both accuracy and precision of this method were calculated for unspiked, low, medium and high QC samples, performing five determinations per level and repeating them for 3 days (Table 6.7).

Table 6.7. Summary of the accuracy and precision results of the LC-QQ-MS method for the analyzed QC samples.

	Within-run accuracy (%nominal value)				Within-run precision (%RSD)				Intermediate precision (%RSD)			
	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC
CIT	93.1	101.6	91.9	98.1	1.3	1.7	1.5	0.8	7.2	10.7	8.5	5.6
CNN	99.0	107.5	94.6	107.4	6.8	1.4	2.9	0.6	10.4	7.9	10.7	4.3
SDMA	98.2	100.1	93.6	101.8	1.7	2.0	1.5	5.4	7.2	7.4	7.8	10.5
nC4	115.0	106.5	88.5	101.3	3.3	3.6	3.7	0.9	11.5	11.6	14.3	10.9
SAM	90.8	109.3	111.7	107.6	6.8	6.4	2.7	3.4	17.7	14.1	13.8	12.5
CIS4DEC	103.7	106.4	99.0	114.0	1.5	1.5	3.7	1.4	13.6	11.4	12.3	12.2
SIP	114.3	102.5	111.5	90.6	12.5	3.3	2.5	3.9	10.4	9.7	15.1	14.3
BIL	98.8	94.9	88.1	114.0	1.4	2.5	6.8	0.6	15.0	13.3	11.8	16.4

Accuracy and precision fell within the acceptable range in this study, as observed in Table 6.7. Indeed, accuracy was between 88.1 % and 115.0 % in all cases, whereas, the relative standard deviation (RSD) of repeated individual measures was below 12.5 % for all the analytes. Likewise, intermediate precision was below 17.7 % for all the compounds.

6.5.3. STABILITY

With the aim of verifying that the quantification is not conditioned by any degradation process that may be suffered by the metabolites either in standard solutions or in plasma, stability of the analytes was assessed. Stability study consisted in the analysis of stock solutions, plasma samples and plasma extracts before and after subjecting them to room temperature, freeze and thaw cycles and freezing them for a defined period of time. Table 6.8 summarizes the results obtained from different stability experiments.

Individual stock solutions proved to be stable frozen for more than a month. Nevertheless, combined stock solutions were less stable. Indeed, these solutions were stable for at least 24 h at room temperature (85–106 %), except for BIL, which rapidly degrades (14 %). In addition, they were stable frozen for 1 month, except for BIL that was stable less than a week (74 %). For that reason, BIL was stored frozen apart from the rest of combined stock solutions and was prepared once a week.

Regarding plasma extracts, they were stable at room temperature for 24 h (85-106 %), and frozen a week (85-112 %), except for CIT (71 %). Besides, no degradation effect was observed neither on plasma extracts after 1 freeze and thaw cycle (94-109 %) nor in plasma samples stored frozen prior to sample treatment (90-107 %).

Table 6.8. Results obtained from stability experiments carried out for the analytes of interest.

COMPOUND	Stock solutions					Plasma samples			Plasma extracts			
	Room temperature 24 h	Frozen 1 week	Frozen 1 month	1 freeze and thaw cycle	2 freeze and thaw cycles	1 freeze and thaw cycle	Room temperature 24 h	Frozen 1 week	1 freeze and thaw cycle	Room temperature 24 h	Frozen 1 week	1 freeze and thaw cycle
CIT	85.6 ± 1.1	103.6 ± 1.7	106.0 ± 0.1	94.1 ± 9.8	113.7 ± 2.5	98.4 ± 7.3	92.7 ± 7.3	71.2 ± 1.2	98.2 ± 0.5	92.7 ± 7.3	71.2 ± 1.2	98.2 ± 0.5
CNN	93.3 ± 8.3	98.4 ± 1.3	98.6 ± 2.9	101.0 ± 9.7	94.5 ± 3.1	101.2 ± 1.2	102.8 ± 1.5	103.4 ± 6.7	106.1 ± 5.4	102.8 ± 1.5	103.4 ± 6.7	106.1 ± 5.4
SDMA	94.2 ± 6.1	96.6 ± 5.5	97.8 ± 0.5	103.1 ± 9.8	97.0 ± 1.6	99.4 ± 2.1	99.7 ± 1.5	93.8 ± 2.5	102.4 ± 4.8	99.7 ± 1.5	93.8 ± 2.5	102.4 ± 4.8
nC4	106.2 ± 0.3	111.2 ± 0.5	90.9 ± 6.9	87.9 ± 5.3	77.4 ± 11.8	106.6 ± 9.9	106.2 ± 2.5	104.5 ± 6.4	103.2 ± 2.0	106.2 ± 2.5	104.5 ± 6.4	103.2 ± 2.0
SAM	97.8 ± 11.3	90.2 ± 14.4	96.7 ± 0.2	92.5 ± 4.8	93.0 ± 11.7	93.9 ± 7.2	96.3 ± 13.1	84.8 ± 6.6	94.0 ± 10.1	96.3 ± 13.1	84.8 ± 6.6	94.0 ± 10.1
CIS4DEC	87.9 ± 0.2	111.3 ± 2.8	69.1 ± 6.5	87.6 ± 6.2	89.3 ± 3.5	95.8 ± 12.2	100.8 ± 2.2	93.5 ± 2.6	99.7 ± 3.4	100.8 ± 2.2	93.5 ± 2.6	99.7 ± 3.4
S1P	94.9 ± 7.0	101.3 ± 13.3	93.3 ± 10.5	85.7 ± 8.1	72.8 ± 17.7	94.9 ± 7.1	89.9 ± 3.9	111.8 ± 14.2	95.9 ± 8.3	89.9 ± 3.9	111.8 ± 14.2	95.9 ± 8.3
BIL	14.0 ± 6.6	74.3 ± 2.5	47.4 ± 14.1	71.5 ± 2.8	29.0 ± 7.5	90.2 ± 7.2	85.1 ± 5.0	109.6 ± 2.8	109.2 ± 14.4	85.1 ± 5.0	109.6 ± 2.8	109.2 ± 14.4

6.6. SAMPLE ANALYSIS

After validating the analytical method, the analysis of real plasma samples was performed. Calibration standards, prepared as explained in section 6.5.1, were analyzed twice, in addition to triplicates of CKD and control extracts which were randomly analyzed. QC samples were analyzed during the batch and quantified to verify the correct performance of the calibration curves obtained.

An example chromatogram of a plasma sample from a CKD patient is showed in Fig. 6.11.

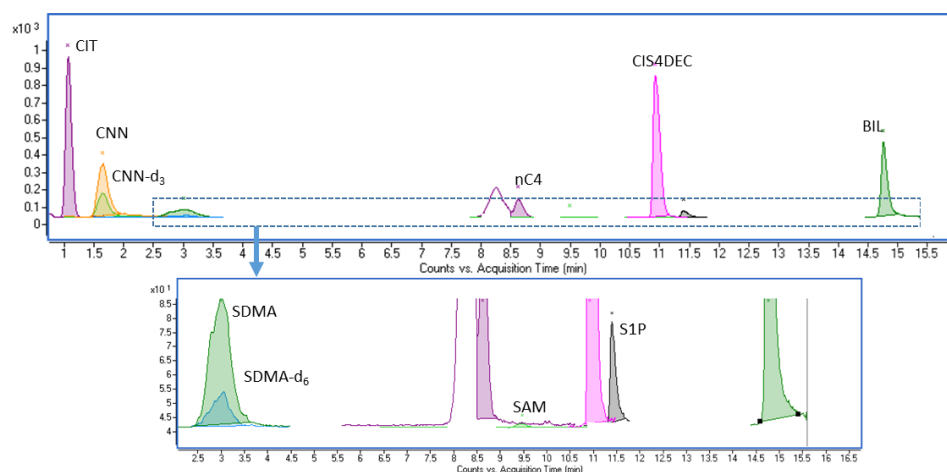


Fig. 6.11. Chromatogram of a plasma sample from a CKD patient.

Samples were quantified dividing the area of the metabolites with the area of internal standards and introducing that ratio into equations of the calibration curve. The following figure shows the results of quantification according to a colour scale for each metabolite and each sample (Fig. 6.12).

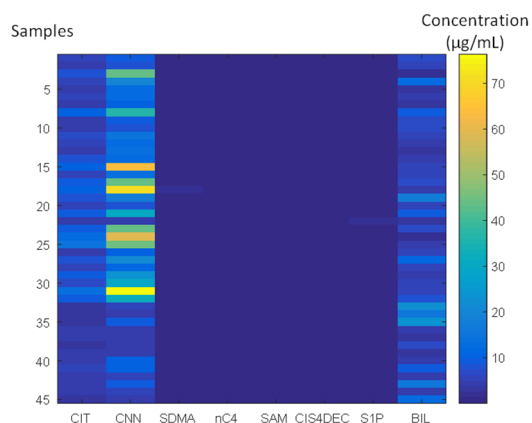


Fig. 6.12. Image graph showing the results of quantification of CKD and control plasma samples.

This figure shows how concentration ranges vary quite a lot between different analytes as well as between samples. For example, between sample CNN concentration differs from around 10 µg/mL to 70 µg/mL. Regarding that SDMA, nC4, SAM, CIS4DEC and S1P showed to be in a lower concentration range, this image was rescaled from 0-70 µg/mL to 0-10 µg/mL (Fig. 6.13).

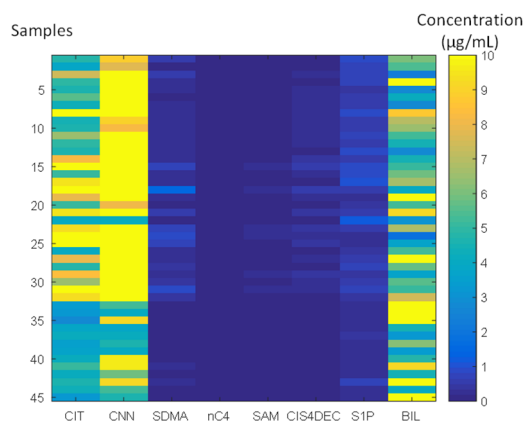


Fig. 6.13. Image graph showing the results of sample analysis rescaled to 0-10 µg/mL.

Between sample concentration differences become patently clear especially for CIT, CNN and BIL. Due to the fact that some other metabolites showed to be in lower concentration ranges, this image was again rescaled from 0-10 µg/mL to 0-1 µg/mL.

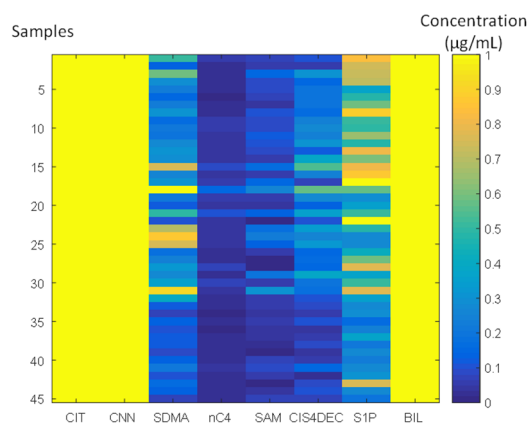


Fig. 6.14. Image graph showing the results of sample analysis rescaled to 0-1 $\mu\text{g/mL}$.

Having a look at all of these figures, it is possible to state three different metabolite groups according to their concentrations: CIT, CNN and BIL containing concentrations from 5 to 70 $\mu\text{g/mL}$, SDMA and S1P containing concentrations from 100 ng/mL to 1 $\mu\text{g/mL}$ and nC4, SAM and CIS4DEC containing lower concentrations.

6.7. DATA ANALYSIS

This newly developed and validated LC-QQQ-MS method is intended for routine analysis in clinical practice. Thus, in principle once established the thresholds for each analyte or the equation in which analyte concentrations will be substituted, there is no need for data analysis. However, regarding that the results obtained from the untargeted metabolomics method in Chapter 5 need to be validated, data analysis of the results obtained from the samples listed in section 6.3.4 has been carried out using Matlab software (Mathworks).

6.7.1. DESCRIPTIVE STATISTICS

With the purpose of having a general idea of the nature of data, data analysis started with descriptive statistics. First, results from quantification of these 8 metabolites were summarized in Table 6.9, where median values for each group and 3rd-97th interquartile range (IQR) are showed.

Table 6.9. Median and IQR obtained from analyte quantification expressed in $\mu\text{g/mL}$.

	Control ($\mu\text{g/mL}$)	CKD ($\mu\text{g/mL}$)
CIT	3.8 (2.9-4.9)	6.5 (4.0-13.5)
CNN	5.2 (3.8-9.6)	15.6 (7.5-71.0)
SDMA	0.115 (0.076-0.193)	0.294 (0.115-0.961)
nC4	0.019 (0.012-0.034)	0.038 (0.017-0.102)
SAM	0.053 (0.031-0.076)	0.083 (0.044-0.269)
CIS4DEC	0.069 (0.034-0.135)	0.195 (0.069-0.543)
S1P	0.275 (0.174-0.617)	0.575 (0.270-1.0)
BIL	6.1 (3.3-23.0)	5.3 (2.2-12.9)

According to this table, CIT, CNN, SDMA, nC4, SAM, CIS4DEC and S1P would be up-regulated for CKD patients, whereas BIL would be down-regulated. The results for amino acids and related compounds match with those obtained in Chapters 4, and the up- and down-regulations of the rest of the compounds with findings from Chapter 5.

Kolmogorov-Smirnov test was used to study distributions of analyte concentrations in control and CKD sample groups. According to this test all the metabolites were under parametric distribution except for BIL.

6.7.2. UNIVARIATE ANALYSIS

The relation between individual analyte concentrations in both sample groups was studied to verify whether this up- and down-regulations observed were significant. Student's *t*-test was applied for parametric variables, whereas *U* test in accordance to Mann Whitney was used to evaluate BIL, the only non-parametric variable.

According to these tests, CIT, CNN, SDMA, nC4, SAM, CIS4DEC and S1P were found to be significantly increased ($p < 0.001$), as expected regarding the results obtained from previous targeted and untargeted studies.

However, for BIL this difference was not found to be significant ($p > 0.05$). Different factors could explain that. On the one hand, the untargeted study was carried out using 24 control samples, whereas for this study only 13 control samples were available and distribution within these samples was wide. On the other hand, the protein precipitation method used for untargeted metabolomics method involved the use of methanol:ethanol (50:50, v/v), which showed to be less appropriate than acetonitrile alone for quantification purposes,

used in this targeted method. Indeed, acetonitrile extraction provided 10 times higher signal of BIL in comparison with the signal obtained from methanol:ethanol extraction, which may imply incomplete extraction of BIL using methanol:ethanol.

This LC-QQQ-MS method validates nC4, CIS4DEC, S1P and BIL, the metabolites found to be significantly altered in CKD patients according to the untargeted metabolomics method, as significant biomarkers. However, regarding that the significance of BIL cannot be corroborated, its use as a potential biomarker should be carefully considered in future analyses.

6.7.3. MULTIVARIATE ANALYSIS

6.7.3.1. Determination of the outlying samples

Regarding that the presence of outlying samples could affect multivariate models built, discarding outliers was considered essential. Determination of outlying samples was performed according to Hierarchical Clustering with single, complete and average linkage (see Fig. 6.15).

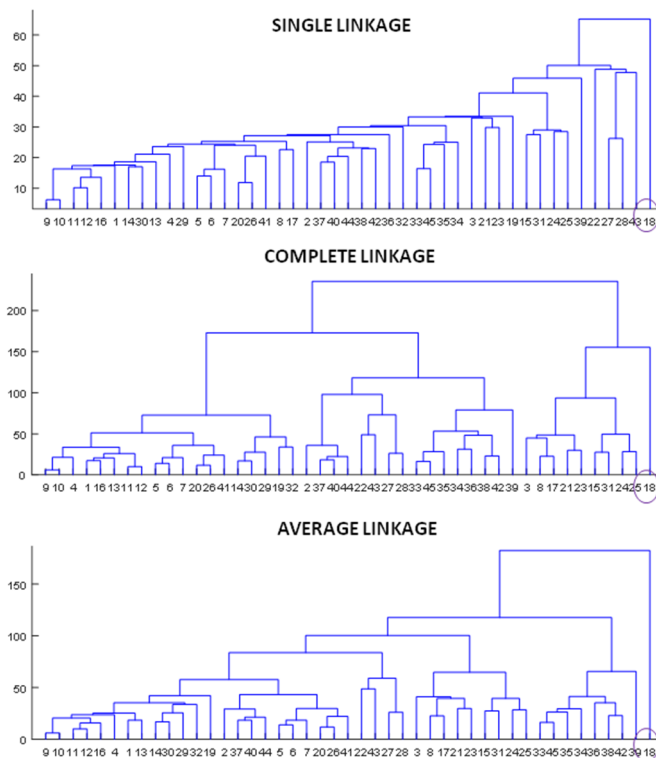


Fig. 6.15. Dendrograms showing Hierarchical clustering according to single, complete and average linkages showing one outlying sample.

As it can be observed from this figure, sample numbered as 18 in this sequence always clustered apart. This sample was tracked and corresponded to the same sample clustered apart in Chapter 4, which also showed a different colour. As a consequence, it was considered an outlier again and discarded from future data analysis.

6.7.3.2. Cross-correlations between metabolites

Potential relations between concentrations of different metabolites were assessed by means of cross-correlations of these analytes by means of SPSS Statistics 23 software (IBM). Table 6.10 summarizes the correlations found between different analytes.

Table 6.10. Cross-correlations found for between metabolites.

CIT	1							
CNN	0.886	1						
SDMA	0.774	0.900	1					
nC4	0.412	0.512	0.658	1				
SAM	0.739	0.858	0.807	0.444	1			
CIS4DEC	0.594	0.617	0.695	0.664	0.618	1		
S1P	0.177	0.252	0.203	0.306	0.108	0.145	1	
BIL	-0.232	-0.214	-0.257	-0.113	-0.245	-0.226	-0.221	1
	CIT	CNN	SDMA	nC4	SAM	CIS4DEC	S1P	BIL

The highest correlations were found for CNN-SDMA ($r=0.900$, $p<0.001$), CNN-CIT (0.886 , $p<0.001$) and CNN-SAM (0.858 , $p<0.001$), matching with the correlations obtained in Chapter 4. Besides, taking into account the interrelations between analytes correlated with CNN, a high correlation was found between SAM-SDMA ($r=0.807$, $p<0.001$). In addition, CIT-SDMA ($r=0.774$, $p<0.001$) and CIT-SAM ($r=0.739$, $p<0.001$) concentrations were somehow related.

However, the analytes added from the untargeted metabolomics discovery (nC4, CIS4DEC, S1P and BIL) did not show high correlations neither with CNN nor with other metabolites from the method. This implies that concentrations of these metabolites are not proportional to those from other analytes, and thus, this could mean that different and independent metabolic pathways might be affected.

6.7.3.3. Scaling and construction of the PCA model

Multivariate data analysis need for scaling and normalization steps to provide meaningful results.

The results obtained in sections 6.6 and 6.7.1 showed that some analytes did not have normal distribution and differences were expected regarding the concentration range of the metabolites. Therefore, logarithm transformation and autoscaling were applied to data matrix to correct for these differences, assuming that all the metabolites are equally eligible to be biomarkers, regardless of their concentration range in plasma.

This scaled and normalized data matrix was used to perform PCA and the resulting biplot, coloured according to CKD and control populations, is showed in Fig. 6.16. Regarding that BIL could not be validated by means of this method, both PCA considering and not considering BIL are showed.

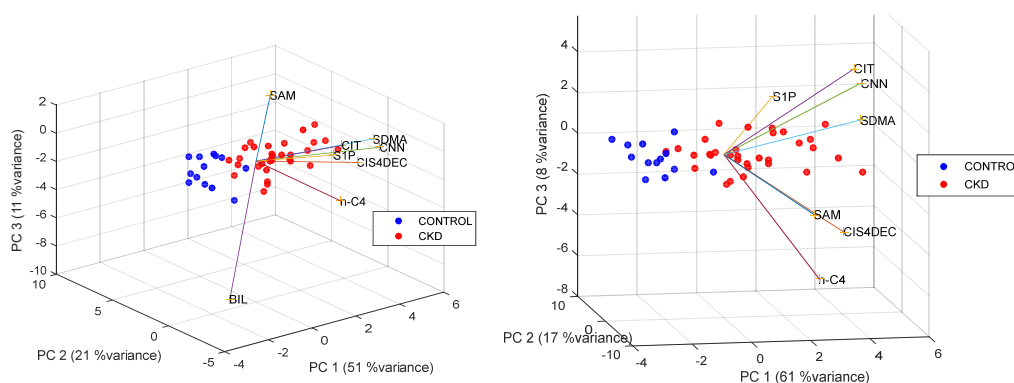


Fig. 6.16. PCA biplot showing separation of CKD and control groups and distribution and distribution of metabolite concentrations, considering (left) and not considering BIL (right).

A separation between groups according to metabolite concentrations is observed in these PCA representations. In addition, it should be noted that the first three principal components accounted for 86 % of variance when leaving out BIL. Moreover, PCA including BIL shows that this metabolite is perpendicular to sample group distribution, which means that probably there is not any relation between BIL concentration and sample distribution.

Besides, regarding that one of the goals of this study is assessing whether these biomarkers would be useful for early diagnosis of CKD, a PCA model has been constructed considering only early CKD samples (those at CKD2 stage) and control samples.

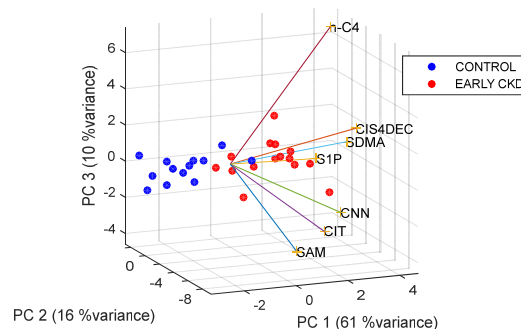


Fig. 6.17. PCA biplot showing separation of early CKD and control patient distribution and of metabolite concentrations, considering all the metabolites except for BIL.

Again, this PCA shows that all the metabolites found to be significant are increased according to the loadings, matching with previous results, and enable a separation of CKD and control paediatrics.

To sum up, this LC-QQQ-MS method developed for routine analysis seems to be able to differentiate between control and CKD paediatrics. Moreover, good separation of control and early CKD samples has been obtained as well. All of the metabolites from targeted and untargeted methods showed to be significant both in univariate and multivariate analysis except for BIL. Besides, BIL did not showed to be significant neither in univariate nor in multivariate analysis and thus future studies in a wider population are needed to study the significance of BIL as a biomarker of CKD in paediatrics.

6.8. CONCLUSIONS

Taking into consideration the proposed goals and the results obtained in this targeted metabolomics study, it can be concluded that:

- ✓ A new ion pairing LC-QQQ-MS analytical methodology has been optimized and validated to analyze significant metabolites found in previous targeted and untargeted metabolomics studies.
- ✓ This method is useful for the validation of the metabolites found to be significant in the untargeted metabolomics study, in addition to including other analytes of interest from the targeted metabolomics study for quantification purposes.
- ✓ Plasma sample treatment was optimized for the analytes of interest and requires a low sample volume and consists in a simple plasma protein precipitation procedure.

- ✓ After optimization and validation, this method was successfully applied to the analysis of both a heterogeneous group of CKD paediatric patients and a control population.
- ✓ The results obtained for the amino acids and related compounds matched well with the results achieved in Chapter 4.
- ✓ Regarding the validation of the results of the metabolites from the untargeted metabolomics method, all the metabolites except for BIL showed to be significant in both univariate and multivariate data analysis, and thus, might be considered potential biomarkers.
- ✓ This method containing heterogeneous analytes enables separation of the samples in CKD-control and early CKD-control groups according to metabolite concentration.
- ✓ Future studies in a wider population would be advisable to study the significance of BIL and the classification performance of this new method.

6.9. **BIBLIOGRAPHY**

1. Gostick D, Kieser B. Mass spectrometry advances. Translational medicine trends in proteomic and small-molecule applications. *Current Trends in Mass Spectrometry*. 2006; 43-46.
2. Naz S, Vallejo M, Garcia A, Barbas C. Method validation strategies involved in non-targeted metabolomics. *J Chromatogr A*. 2014; 1353: 99-105.
3. Martin J-C, Maillot M, Mazerolles G, Verdu A, Lyan B, Migne C, et al. Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. *Metabolomics*. 2015; 11: 807-821.
4. Nagana Gowda GA, Raftery D. Biomarker discovery and translation in metabolomics. *Curr Metabolomics*. 2013; 1: 227-240.
5. Benito S, Sanchez A, Unceta N, Andrade F, Aldamiz-Echevarria L, Goicolea MA, et al. LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease. *Anal Bioanal Chem*. 2016; 408: 747-760.
6. Benito S, Sanchez-Ortega A, Unceta N, Jansen J, Postma G, Andrade F, et al. Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data. *J Pharm Biomed Anal*. 2017; doi: <https://doi.org/10.1016/j.jpba.2017.10.036> [In press]
7. Rosenberg M, Kalda R, Kasiulevicius V, Lember M. Management of chronic kidney disease in primary health care: position paper of the European Forum for Primary Care. *Qual Prim Care*. 2008; 16: 279-294.

8. ACD/Labs. Advanced Chemistry Development Software V11.02
9. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens J-P, Morla A, et al. ESI-MS/MS analysis of underivatized amino acids: A new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionization mode. *Rapid Commun Mass Sp.* 2003; 17: 1297-1311.
10. Oosterink JE, Buijs N, van Goudoever JB, Schierbeek H. A novel method for simultaneous measurement of concentration and enrichment of NO synthesis-specific amino acids in human plasma using stable isotopes and LC/MS ion trap analysis. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2014; 958: 10-15.
11. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens J-P, Bouchu D. Ion-pairing reversed-phase liquid chromatography/electrospray ionization mass spectrometric analysis of 76 underivatized amino acids of biological interest: A new tool for the diagnosis of inherited disorders of amino acid metabolism. *Rapid Commun Mass Sp.* 2005; 19: 1587-1602.
12. Martens-Lobenhoffer J, Bode-Boeger SM. Quantification of L-arginine, asymmetric dimethylarginine and symmetric dimethylarginine in human plasma: A step improvement in precision by stable isotope dilution mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2012; 904: 140-143.
13. Bishop MJ, Crow B, Norton D, Paliakov E, George J, Bralley JA. Direct analysis of un-derivatized asymmetric dimethylarginine (ADMA) and L-arginine from plasma using mixed-mode ion-exchange liquid chromatography-tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2007; 859: 164-169.
14. Zotti M, Schiavone S, Tricarico F, Colaianna M, D'Apolito O, Paglia G, et al. Determination of dimethylarginine levels in rats using HILIC-MS/MS: an in vivo microdialysis study. *J Sep Sci.* 2008; 31: 2511-2515.
15. 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Official Journal of the European Communities.* 2002.
16. The Merck Index, Bilirubin (entry# 1235). 11th ed.
17. Dawson RMC. *Data for Biochemical Research.* 3rd ed: Clarendon; 1986; p. 220-221.
18. Kirsch SH, Knapp J-P, Geisel J, Herrmann W, Obeid R. Simultaneous quantification of S-adenosyl methionine and S-adenosyl homocysteine in human plasma by stable-isotope dilution ultra performance liquid chromatography tandem mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci.* 2009; 877: 3865-3870.



CHAPTER VII.

General conclusions

Taking into account the main objective of the present study about new biomarker discovery for the early diagnosis and progression monitoring of CKD in paediatrics using different metabolomics approaches, it could be concluded that:

- ✓ As far as **targeted metabolomics approach**, a new LC-QTOF-MS analytical methodology has been optimized for the analysis of 16 amino acids, amino acid derivatives, and related compounds from the arginine-creatine pathway, the arginine methylation, and the urea cycle in a single run, in MS and MS/MS mode in less than 19 min.

This validated method was successfully applied to the analysis of a heterogeneous group of CKD pediatric patients and an age-related control population. Univariate analysis of these results found significant differences for glycine, citrulline, creatinine, asymmetric dimethylarginine, symmetric dimethylarginine and dimethylglycine.

Moreover, three new metabolites, citrulline, S-adenosylmethionine and symmetric dimethylarginine, have been proposed as potential biomarkers in addition to the commonly used creatinine to be implemented since they enable a better diagnosis of early stages of CKD in paediatrics. Furthermore, these 4 analytes showed greater improvement of the diagnosis regardless of the stage of CKD, and a relation between concentration and CKD stage was found.

- ✓ Regarding **untargeted metabolomics approach**, a LC-QTOF-MS analytical method has been optimized with the aim of extracting as many features as possible and to obtain significant features to differentiate between control and CKD paediatrics.

Five features were found to be significantly up- or down-regulated with high confidence in CKD paediatrics, as they matched in the two different chemometric approaches carried out. Moreover, classification of control and early CKD samples using these features showed an overall accuracy of 97 %.

Identification of these 4 out of 5 entities was fully achieved, with 1st level compound identification confidence according to Metabolomics Standards Initiative as: *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine, bilirubin and sphingosine-1-phosphate.

- ✓ Finally, a **LC-QQQ-MS analytical method** has been developed and validated for the biomarkers discovered within the previous two approaches: citrulline, S-adenosylmethionine, symmetric dimethylarginine, creatinine, *n*-butyrylcarnitine, *cis*-4-decenoylcarnitine, bilirubin and sphingosine-1-phosphate. Therefore, this method enables simultaneous quantification of all of these metabolites in routine analysis.

This LC-QQQ-MS method complies with the requirement of untargeted metabolomics studies for biomarker validation by means of analytical method development aimed at targeting the metabolites of interest. After univariate and multivariate analyses, all the metabolites except for bilirubin showed to be significant and thus, could be considered potential biomarkers.

Future studies using a wider population would be of interest to be able to study the significance of bilirubin, as well as the classification performance of this new analytical method.



ANNEX.

Articles published in scientific journals

ANNEX I:

LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease



LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease

Sandra Benito¹ · Alicia Sánchez² · Nora Unceta¹ · Fernando Andrade³ · Luis Aldámiz-Echevarria³ · M. Aránzazu Goicolea¹ · Ramón J. Barrio¹

Received: 29 July 2015 / Revised: 15 October 2015 / Accepted: 27 October 2015 / Published online: 4 November 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Chronic kidney disease (CKD) is a major epidemiologic problem which causes several disturbances in adults and in pediatrics. Despite being a worldwide public health problem, information available for CKD in the pediatric population is scarce. For that reason, an ion-pair reversed-phase liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-QTOF-MS) method has been developed and validated in order to analyze 16 amino acids, amino acid derivatives, and analogous compounds related to the arginine-creatine metabolic pathway that are suspicious of being increased or decreased in plasma from patients with CKD. The analytical method involved the addition of dithiothreitol, a reducing agent which reduces disulfide and thus giving total aminothiol concentration, as well as a simple precipitation of plasma proteins. Moreover, despite amino acids being usually derivatized to improve their retention time and to enhance their signal, for this method, an ion-pairing reagent was used,

thus avoiding the need for derivatization. Subsequently, analysis of plasma from pediatric patients suffering from CKD and control pediatrics was carried out. As a result, glycine, citrulline, creatinine, asymmetric dimethylarginine (ADMA), and symmetric dimethylarginine (SDMA) were significantly increased in patients with CKD, regardless of their creatinine level, whereas in addition to these compounds dimethylglycine was also increased when CKD patients had plasma creatinine concentrations above $12 \mu\text{g mL}^{-1}$, thus all are suggested as potential biomarkers for renal impairment.

Keywords Chronic kidney disease (CKD) · Arginine-creatine metabolic pathway · Biomarkers · Plasma · LC-QTOF-MS

Introduction

Chronic kidney disease (CKD) is a major epidemiologic problem in which renal hypofunction occurs due to an irreversible kidney damage that can further progress. This situation causes the disturbance of electrolyte balance as well as the accumulation of waste products which can lead to other health problems, such as cardiovascular disease [1, 2]. In children, growth retardation usually occurs due to the following factors related to CKD as well: protein and calorie malnutrition, metabolic acidosis, end-organ growth hormone resistance, and renal bone disease [3]. Moreover, even if the occurring disturbances such as acidosis, anemia, vitamin D, and serum calcium and phosphorus levels are solved, a great amount of children with CKD continue to grow poorly, being catch-up growth only exceptionally seen [3, 4]. Therefore, it is essential to start treatment at the right point in the progression of CKD to prevent adverse outcomes [5].

Electronic supplementary material The online version of this article (doi:10.1007/s00216-015-9153-9) contains supplementary material, which is available to authorized users.

✉ Ramón J. Barrio
r.barrio@ehu.es

¹ Department of Analytical Chemistry, Faculty of Pharmacy, University of the Basque Country (UPV/EHU), Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain

² Central Service of Analysis (SGiker), University of the Basque Country (UPV/EHU), Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain

³ Group of Metabolism, BioCruces Health Research Institute, CIBER de Enfermedades Raras (CIBERER), Plaza de Cruces 12, 48903 Barakaldo, Spain

Despite that different measures are taken, in most pediatric CKD patients, end-stage renal disease (ESRD) inevitably occurs [3]. ESRD is a condition in which patients need for renal replacement treatment (RRT). Hemodialysis (HD), peritoneal dialysis (PD), and renal transplant (Tx) are the three modalities of RRT currently in use [6]. Moreover, according to the registries collected by the European Society for Pediatric Nephrology (ESPN) and the European Renal Association and European Dialysis and Transplantation Association (ERA-EDTA) between 2009 and 2011, the prevalence of children receiving renal replacement therapy (RRT) was 27.9 per million age-related population (pmarp) in patients aged 0–14 years, taking into account 37 European countries. In addition, mortality in pediatric patients being treated with RRT is 55-fold higher than in general pediatric population [7].

Furthermore, CKD is difficult to diagnose, to follow in progression, and to evaluate the response to therapy [8]. In the clinical practice, the diagnosis is traditionally based on pathology test results as well as on etiology [5]. Usually, serum creatinine concentration (SCr) for the assessment of glomerular filtration rate (GFR), blood urea nitrogen level (BUN), and urinalysis are the evidences of election to detect kidney disease [9]. Nevertheless, these biomarkers are not appropriate to detect CKD in early stages [10–12], and CKD is under-recognized in the general population as serum creatinine alone is not an accurate biomarker to measure the kidney function [13]. Besides, even if CKD is a worldwide public health problem and extensive research is being done in adult population, little information is available for CKD in the pediatric population [1].

With the aim of finding the ideal marker to make an early diagnose of CKD and follow both the progression of the disease and the response to the treatment, a new analytical methodology has been developed for the determination of 16 amino acids, amino acid derivatives, and related compounds which are thought to be decreased or increased in pediatric CKD patients: L-homoarginine, L-homocysteine, L-arginine, symmetric dimethylarginine, asymmetric dimethylarginine, *N,N*-dimethylglycine, *S*-adenosylhomocysteine, *S*-adenosylmethionine, citrulline, betaine, creatine, creatinine, glutathione, methionine, glycine, and cysteine. All of them are related to the arginine-creatine metabolic pathway or other arginine-related metabolic pathways such as the urea cycle or the arginine methylation (Fig. 1) and could be increased or decreased in CKD as shown in the Electronic Supplementary Material (ESM) in Table S1.

Nowadays, there is not any analytical method which gathered all these compounds together and enables their quantification. Amino acid analysis by MS detection is challenging since they are small and polar compounds, and therefore, it is difficult to resolve them by conventional reverse phase HPLC and interferences occurring in the low mass range suppress the MS sensitivity [14].

Moreover, they do not have functional groups suitable for UV or fluorescence detection, and because of that, amino acids have been traditionally analyzed and preceded by derivatization with different compounds such as 9-fluorenylmethyl chloroformate (FMOC), *o*-phthalaldehyde (OPA), phenylisothiocyanate (PITC), or 5-dimethylamino-1-naphthalenesulphonyl-chloride (DANSYL), which react with the amine group, or with butanol, which reacts with the carboxylic group [15, 16]. The amino acid derivatization not only improves their detection but also their separation [16]. Some of these derivatization reagents have been widely used for LC/MS, such as OPA [17, 18], FMOC [16, 19], and butanol [16].

Nevertheless, the trend is towards avoiding derivatization, and for that reason, the use of ion-pairing agent appears to be useful to improve the retention of these compounds in liquid chromatography. There are several methods developed for amino acid analysis using perfluoroheptanoic acid (PFHA), also known as tridecafluoroheptanoic acid (TDFHA), such as the LC/MS/MS method reported by Piraud et al. [15, 20] and Qu et al. [21] or another similar described by Armstrong et al. using a HPLC-TOF/MS method [22].

When using mass spectrometric detection without derivatization, the use of triple quadrupole mass spectrometer [15, 20] and time-of-flight mass spectrometer [22] has been reported. The first one allows the identification and quantification of amino acids by means of tandem mass spectrometry (MS/MS), whereas the second one has the ability of providing accurate mass of each analyte.

Current method uses ion-pairing liquid chromatography quadrupole time-of-flight mass spectrometer (LC-QTOF), allowing the improvement in the retention of some of the amino acids with no need for derivatization. To the best of our knowledge, it is the first time that LC-QTOF has been used for this aim guaranteeing accurate mass and the achievement of MS and MS/MS spectra at the same time. Indeed, this enables to distinguish ADMA and SDMA isomers, which have identical ionization in MS mode and elute at the same retention time, as their MS/MS monitorization provides unique and specific transitions.

However, thiol groups in some aminothiols compounds such as glutathione, cysteine, and homocysteine are critical intracellular and extracellular redox buffers, which form disulfides (–S–S–) when undergo oxidative coupling reactions, thus binding to proteins or to other aminothiols [23]. When measuring free aminothiol concentration, oxidation and reduction reactions could occur, leading into under- or overestimation in quantification, unless thiol groups are blocked by derivatization with different compounds at the time of blood extraction [24]. Therefore, in order to obtain

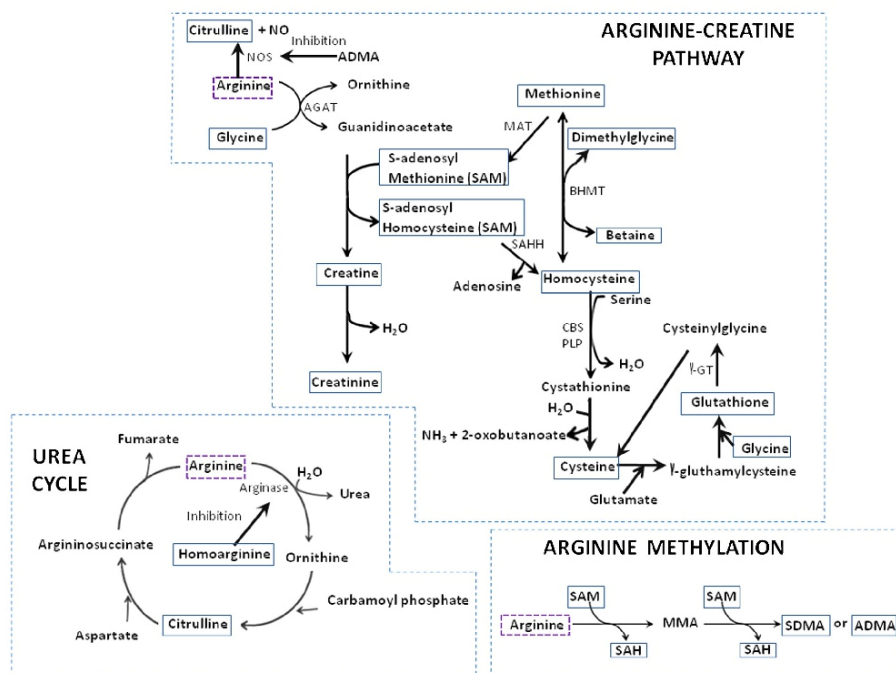


Fig. 1 Relationship between the arginine-creatine pathway, arginine methylation cycle and urea cycle. *AGAT* arginine:glycine amidinotransferase, *NOS* nitric oxide synthase, *MAT* methionine adenosyltransferase, *SAHH* SAH

hydrolase, *BHMT* betaine-homocysteine S-methyltransferase, *PLP* pyridoxal phosphate, *CBS* cystathionine beta synthase, γ -*GT* gamma glutamyl transpeptidase

their total concentration in plasma, dithiothreitol has been used as a reducing agent [25–27].

Material and methods

Chemical and reagents

Acetonitrile used in the mobile phase and LC-MS grade ammonium formate eluent additive were obtained from Scharlau (Sentmenat, Spain) and Fluka Analytical, Sigma-Aldrich (Steinheim, Germany), respectively. Perfluoroheptanoic acid used as ion-pairing reagent was acquired from Acros Organics (New Jersey, USA). The mobile phase was filtered prior to use through 0.1 μ m filters from Millipore Omnipore (Watford, Ireland).

For standard preparation, in addition to acetonitrile supplied by Scharlau, LC-MS grade methanol from the same manufacturer, ultra-high purity water obtained from tap water

pre-treated by Elix reverse osmosis, and subsequent filtration by a Milli-Q system from Millipore (Bedford, MA, USA) and chlorhydric acid obtained from Merck (Darmstadt, Germany) were used.

Standards were obtained from different manufacturers. L-Methionine, glycine, L-arginine, L-homocysteine, N^G, N^G -dimethyl-L-arginine di(*p*-hydroxyazobenzene-*p*'-sulfonate) or SDMA, N^G, N^G -dimethylarginine dihydrochloride (ADMA), S-adenosyl-L-methionine (SAM), S-adenosyl-L-homocysteine (SAH), and citrulline were provided by Sigma-Aldrich (Steinheim, Germany). Betaine, reduced glutathione (GSH), L-cysteine, and creatine were supplied by TCI (Tokyo, Japan). Creatinine was purchased from Alfa Aesar (Karlsruhe, Germany). Finally, *N,N*-dimethylglycine and homoarginine were obtained from Fluorochem (Hadfield, UK).

In addition, the following isotopically labeled compounds were used as internal standard: creatine- d_3 H₂O (methyl- d_3) from CDN Isotopes (Quebec, Canada), N^G, N^G -dimethyl-L-arginine- d_6 , and creatinine- d_3 supplied by Toronto Research

Chemicals, TRC-Canada (North York, Canada), and glutathione- $^{13}\text{C}_2^{15}\text{N}$ supplied by Cambridge Isotope Laboratories (Andover, MA, USA).

As a surrogate matrix, Serasub[®] was obtained from the manufacturer CST Technologies (Great Neck, NY, EEUU). Serasub[®] is a liquid, protein-free synthetic polymer in buffered solution (pH 7.4) equivalent to serum and plasma in terms of specific gravity, viscosity, and osmolality [28].

The thiol reductant agents, dithiothreitol and 2-mercaptoethanol, were obtained from Fisher Scientific (Pittsburgh, PA, USA) and Sigma-Aldrich (Steinheim, Germany), respectively.

Two different solid phase extraction cartridges were also assayed: Oasis MCX 3 cc, 60 mg supplied by Waters (Milford, MA, USA), and Hybrid SPE Phospholipid Ultra 30 mg/1 mL from Supelco (Bellefonte, PA, USA).

LC-QTOF conditions

Chromatographic analysis was accomplished on an Agilent 1200 Series HPLC system coupled to a hybrid quadrupole time-of-flight mass spectrometer Agilent 6530 Series from Agilent Technologies (Santa Clara, CA, USA). Data acquisition was done by Agilent MassHunter version B.05.01 software.

Chromatographic separation was studied on three different columns: Gemini-NX C18 110 Å (4.6×150 mm, 5 µm) column from Phenomenex (Torrance, CA, USA), as well as Zorbax Eclipse plus C8 80 Å (2.1×150 mm, 5 µm) and Poroshell EC-C18 120 Å (4.6×100 mm, 2.7 µm) both from Agilent Technologies (Santa Clara, CA, USA). Besides, once selected, the Poroshell column rapid resolution LC inline filter (4.6 mm, 0.2 µm) from the same manufacturer was added to the column. The column was kept at 30±0.8 °C. A binary solvent system consisting of 0.5 mM PFHA and 5 mM ammonium formate (A) and acetonitrile (B) was used for the elution gradient. The following gradient program was used: 0–2 min, 5% eluent B, 2–10 min linear gradient increase from 5 to 20% eluent B, 10–11 min linear gradient increase from 20 to 60% eluent B, 60% eluent B was held for 2 min and the gradient was decreased from 60 to 5% in 2 min time, and finally 5% eluent B was held for 2 min, with a post-time of 2 min. A flow rate of 400 µL min⁻¹ was selected.

Two microliters of plasma extract were injected in the autosampler, and in order to prevent carry over, the needle was rinsed for 120 s using a flushport with isopropanol/water (50:50, v/v). In addition, injector valve was switched 3 times during each analysis run from bypass to mainpass and vice versa.

Agilent Jet Stream ESI parameters were optimized in order to get maximum signal intensities and were set as follows: dry gas flow, 10 L min⁻¹; dry gas temperature, 300 °C; nebulizer

pressure, 40 psig; sheath gas temperature, 325 °C; and sheath gas flow, 11 L min⁻¹.

The MS detector operated in low mass range (1700 *m/z*), 2 GHz extended dynamic range, and centroid mode was used for data collection and storage. Concerning instrument resolution, it gave a 12,000 full width at half maximum typical resolution (FWHM). In order to guarantee mass accuracy, a reference solution was directly infused into the source, enabling continuous internal calibration during analysis and ensuring accuracy and reproducibility. For that purpose, purine (121.0509 *m/z*) and HP-921 (922.0098 *m/z*) signals were used.

Moreover, QTOF allows analysis in MS or MS/MS mode. When working in MS mode, total transmission ion mode (TTI) is used, that is, all of the ions are conducted through the quadrupole without any isolation or fragmentation, which results in a MS spectrum that contains the exact mass of compounds. However, in MS/MS mode, the selected precursor ion is isolated in the quadrupole and subsequently fragmented in the collision cell, yielding characteristic MS/MS spectra. MS spectra were acquired with a scan rate of 2.2 spectra s⁻¹ in MS mode and with a scan rate of 3 spectra s⁻¹ in MS/MS mode. With the aim of obtaining characteristic and abundant product ions in MS/MS spectra, nitrogen gas in the collision cell and ramping energies from 10 to 35 V were assayed. MS spectra acquisition was carried out in a mass range from 50 to 1000 *m/z* whereas MS/MS data acquisition was obtained from 30 to 1000 *m/z*.

The retention time of each analyte in plasma, accurate mass measurement of the precursor ion with an error less than 5 ppm (MS spectra), and accurate mass measurement of the selected product ion (MS/MS spectra) were the necessary mainstays for the positive identification of the target compounds. Regarding the EU's decision 2002/657/CE proposed for the collection of identification points, confirmation of the presence of the analytes in samples is possible by using the developed methodology [29].

Sampling

Thirty-two patients suffering from chronic kidney disease aged 3–17 years and 24 patients not suffering from chronic kidney disease aged 6–18 years were recruited for the study. From the total of patients with CKD, 14 were in stage 2 (GFR 60–89 mL/min/1.73 m²), six patients in stage 3 (GFR 30–59 mL/min/1.73 m²), seven patients in stage 4 (GFR 15–29 mL/min/1.73 m²), and five patients in stage 5 (GFR < 15 mL/min/1.73 m²). Twelve of the patients suffering from CKD received RRT: 4 of them were treated by hemodialysis and another eight patients received kidney transplant.

All CKD patients met the following inclusion criteria: subjects were followed up in Cruces University Hospital and were clinically stable at the time of the study. Patients were discarded when meeting any of the following exclusion

criteria: hepatopathy, anuria, and/or insulin-dependent diabetes mellitus. Control patients were selected for being healthy children, aged 6–19 years, who had a minor surgery in Cruces University Hospital. Characteristics of the patients enrolled in the study are shown in Table 1.

Blood samples were taken in the morning after an overnight fasting. Blood samples were cooled in an ice-water bath and immediately centrifuged at 1000g for 5 min at 4 °C. Samples were stored at –80 °C until treatment and analysis was carried out.

The study protocol was approved by the Ethics Committee of Clinic Research of Cruces Hospital, and informed consent was given by patients' parents.

Sample treatment

First of all, samples were thawed and 50 μL of plasma were placed in an Eppendorf tube. Then, the addition of the internal standards was made by spiking 10 μL of a mix solution containing 15 $\mu\text{g mL}^{-1}$ of creatinine- d_3 , creatine- d_3 , and glutathione- $^{13}\text{C}_2^{15}\text{N}$ and 5 $\mu\text{g mL}^{-1}$ of SDMA- d_6 . After that, the reduction of aminothiols compounds was carried out by adding 50 μL of dithiothreitol (77,000 mg L^{-1}) and incubating for 15 min at room temperature. Then, plasma proteins were precipitated by adding 150 μL of cold acetonitrile and the mixture was vortexed before centrifuging it at 13,000 rpm for 10 min at 4 °C. The supernatant was transferred to a chromatography vial and evaporated in nitrogen stream. Finally, reconstitution in 200 μL of ammonium formate 5 mM was performed. Each sample was extracted in triplicate.

Standards and QC samples

Stock standard solutions containing around 1000 mg L^{-1} of betaine, creatine, citrulline, *N,N*-dimethylglycine, homoarginine, creatinine, SAH, homocysteine, SAM, and ADMA were prepared in 0.01 N HCl, with the aim of contributing to their stability. On the other hand, methionine, glycine, arginine, cysteine, and glutathione stock standard solutions were prepared in water. Finally, SDMA was dissolved in methanol. Concerning internal standards, creatinine- d_3 and SDMA- d_6 were prepared in methanol, whereas creatine- d_3 and glutathione- $^{13}\text{C}_2^{15}\text{N}$ were prepared in water. These solutions were stored frozen at –40 °C. Intermediate mixed

dilutions were prepared weekly in ammonium formate 5 mM to obtain working solutions at different concentrations.

Calibration standards were prepared in pooled plasma made of both CKD samples and control samples, which had been previously quantified by standard additions. Pooled plasma was doped to yield concentrations ranging between 159 ng mL^{-1} and 1.4 $\mu\text{g mL}^{-1}$ for ADMA, 14.3 and 46.8 $\mu\text{g mL}^{-1}$ for arginine, 3.1 and 20.6 $\mu\text{g mL}^{-1}$ for betaine, 2.8 and 17.8 $\mu\text{g mL}^{-1}$ for citrulline, 9.7 and 42.2 $\mu\text{g mL}^{-1}$ for creatine, 15.2 and 30.2 $\mu\text{g mL}^{-1}$ for creatinine, 13.6 and 111.1 $\mu\text{g mL}^{-1}$ for cysteine, 1 and 17.3 $\mu\text{g mL}^{-1}$ for dimethylglycine, 1.8 and 10.5 $\mu\text{g mL}^{-1}$ for glutathione, 3.2 and 19.5 $\mu\text{g mL}^{-1}$ for glycine, 462 ng mL^{-1} and 3.7 $\mu\text{g mL}^{-1}$ for homoarginine, 0.5 and 8.7 $\mu\text{g mL}^{-1}$ for homocysteine, 2.7 and 11.5 $\mu\text{g mL}^{-1}$ for methionine, 36 and 524 ng mL^{-1} for SAH, 55 and 555 ng mL^{-1} for SAM, and finally, 291 ng mL^{-1} and 1.5 $\mu\text{g mL}^{-1}$ for SDMA. After being doped, both calibration standards and real samples underwent the same sample treatment.

Furthermore, with the aim of correcting the signal of the analytes of interest, isotopically labeled compounds were added in the quantity previously reported for samples. Despite using isotopically labeled compounds for every analyte would be the ideal option; it is economically unfeasible. For that reason, 4 different isotopically labeled compounds were used, according to their retention time and chemical properties to correct both the recovery of the extraction procedure as well as the signal suppression caused by co-eluting compounds: glutathione- $^{13}\text{C}_2^{15}\text{N}$ for glutathione; creatine- d_3 for glycine, cysteine, homocysteine, citrulline, dimethylglycine, betaine, and creatine; creatinine- d_3 for methionine, creatinine, arginine, and homoarginine; and SDMA- d_6 for SDMA, ADMA, SAH, and SAM.

For method validation, quality control samples (QC) were prepared at four levels of concentration: unspiked QC, low QC, medium QC, and high QC. The unspiked QC was previously quantified by standard additions. The high QC samples, consisted of the same pool but spiked with 12.5 $\mu\text{g mL}^{-1}$ for glutathione, glycine, dimethylglycine, betaine, creatine, methionine, and creatinine, with 75 $\mu\text{g mL}^{-1}$ for cysteine, with 2.5 $\mu\text{g mL}^{-1}$ for homoarginine and homocysteine, with 0.4 $\mu\text{g mL}^{-1}$ for ADMA and SDMA, and with 0.25 $\mu\text{g mL}^{-1}$ for SAH and SAH. Concerning low and medium QC samples, same pooled plasma was spiked with the 10 and 50 % of the concentration spiked for each analyte in the high QC sample.

Regarding sample quantification, calibration curve was analyzed, sample analysis was performed randomizing both CKD and control samples, and unspiked and medium QC samples were included between samples.

Table 1 Characteristics of the patients enrolled in the study

	CKD	Controls
Age (median, interquartil range)	10.8 (3–17.1)	10 (6–17.3)
Gender (M/F)	17/15	18/6

Results and discussion

Optimization of LC-QTOF parameters

For the efficient separation and subsequent detection of the analytes, different mobile phases and columns were tested. The chromatographic separation of amino acids and amino acid derivatives frequently involves the need of derivatization, as they hardly retain because of their high polarity. One alternative for derivatization is the use of ion-pairing liquid chromatography. For that purpose, perfluoroheptanoic acid (PFHA) is one of the most commonly used ion pairing reagents and the majority of the authors report the use of a concentration of 0.5 mM [15, 20, 22, 30, 31]. As the use of 0.5 mM PFHA acidifies the aqueous phase (pH 3.2), in order to obtain a less acidic solution, the addition of 5 mM ammonium formate was considered and selected (eluent A), obtaining a pH 4.3 for the aqueous phase. Taking into account the pKa value of the analytes, the pH obtained by using ammonium formate 5 mM would improve the stability and enhance ionization capability of the compounds, as for ESI interface pH of the mobile phase is of undoubted importance. Eluent A was stirred continuously in order to keep it homogeneous and guarantee repetitive retention times. For the organic solution, acetonitrile was chosen (eluent B). The addition of the PFHA to the organic solution was also tested, but because no improvement was obtained with its addition, the use of ion-pairing reagent in the organic solution was discarded.

Initially, 3 different reverse-phase columns were tested: Gemini-NX C18 110 Å (4.6 × 150 mm, 5 μm), Zorbax Eclipse plus C8 (2.1 × 150 mm, 5 μm), and Poroshell EC-C18 120 Å (4.6 × 100 mm, 2.7 μm). Preliminary studies showed that the last one performed better in terms of resolution and retention of the analytes of interest in comparison with the previous columns. The effect of different column temperatures was also assayed by maintaining it at 30, 35, and 40 °C. No differences were found between different analytes in terms of resolution, so a temperature of 30 °C was chosen.

Taking into account the characteristics of the compounds of interest and the preliminary studies carried out, a flow rate of 400 μL min⁻¹ was selected. Regarding the gradient program, it was optimized to achieve a reasonable separation and retention of analytes of interest on column and was set as previously mentioned in *LC-QTOF conditions*. A representative LC-QTOF chromatogram combining MS and MS/MS is shown in Fig. 2.

Despite the use of ion-pairing reagent that is usually related to retention time drift, the proposed method did not have a significant retention time drift with no need of using 100 % acetonitrile during the analysis as reported in previous analytical methods [22], being the relative standard deviation (RSD) for the retention time below 1.1 % within-run and 2.9 %

between-run for all the analytes ($n=5$), after the injection of one standard solution or sample before starting the run.

Because of the high polarity of the analytes, both conventional ESI interface and an electrospray interface with Agilent Jet Stream[®] thermal gradient technology were tested. Agilent Jet Stream[®] technology usually enhances signal of analytes with a lower RSD at the limit of quantification by both reducing ion dispersion at normal flow rate and decreasing the amount of neutral solvent clusters, as an auxiliary nitrogen gas with concentric orientation towards to the nebulizer spray is used in this interface. As expected, Agilent Jet Stream[®] ESI interface resulted in a higher sensitivity for nearly all the compounds. Despite the varying nature of the tested compounds, positive ion mode was selected for being appropriate for all of the compounds.

Working conditions for the ESI Jet Stream[®] interface were optimized to maximize the intensity of the signals. Sheath gas temperatures of 300, 325, and 350 °C were assayed, as well as sheath gas flows of 10 and 11 L min⁻¹. Similarly, gas temperatures of 300, 325, and 350 °C and the use of a drying gas flow of 10 and 11 L min⁻¹ were also tested. Finally, the conditions previously reported in *LC-QTOF conditions* section were selected for being the best, taking into account all the analytes of interest.

In addition, fragmentor voltages ranging from 75 to 125 V were tested, being selected the 75 V as the most adequate for all the compounds and especially for dimethylglycine since for the rest of the voltages, it was not detected. Nozzle voltages ranging from 0 to 1000 V were assayed as well, finally deciding not to apply any nozzle voltage during the analysis as it did not enhance analyte signal for those compounds with low signal to noise ratio.

All the analytes ionize as the protonated form $[M+H]^+$. In this method, as no derivatization reaction nor specific column was used, MS/MS was the only way to distinguish ADMA and SDMA isomers, as they ionize in the same way in MS mode and they also elute at the same retention time. For that reason, in order to obtain specific fragments, different collision energies (CE) were assayed and it was concluded that a CE of 10 V permitted the monitorization of unique and specific transitions: 203.1503 > 46.0657 for ADMA and 203.1503 > 172.1086 for SDMA (Table 2).

Sample treatment

Initially, two different sample treatment options were evaluated in order to achieve a good extraction and clean-up of the sample: protein precipitation (PPT) and solid-phase extraction (SPE).

Regarding protein precipitation, acetonitrile at -40 °C was used for better stability of the analytes. After protein precipitation, different SPE cartridges were assayed with the aim of

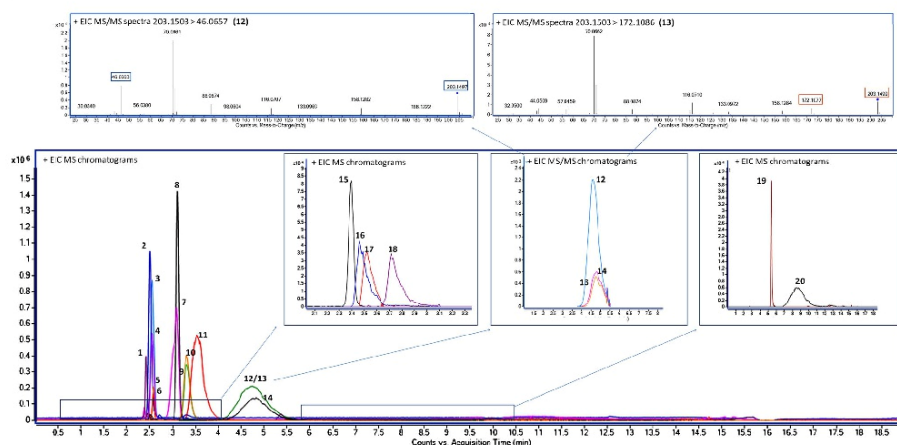


Fig. 2 MS/MS spectra of ADMA and SDMA and chromatogram corresponding to a working standard solution in both MS and MS/MS mode. 1: citrulline 1.5 $\mu\text{g mL}^{-1}$; 2: betaine 0.75 $\mu\text{g mL}^{-1}$; 3: creatine 1.5 $\mu\text{g mL}^{-1}$; 4: creatine- d_3 0.75 $\mu\text{g mL}^{-1}$; 5: glutathione 9 $\mu\text{g mL}^{-1}$; 6: glutathione- $^{13}\text{C}_2^{15}\text{N}$ 7.5 $\mu\text{g mL}^{-1}$; 7: arginine 1.5 $\mu\text{g mL}^{-1}$; 8: methionine 9 $\mu\text{g mL}^{-1}$; 9: creatinine 0.75 $\mu\text{g mL}^{-1}$; 10: creatinine- d_3 0.75 $\mu\text{g mL}^{-1}$; 11: homoarginine 3 $\mu\text{g mL}^{-1}$; 12: ADMA 0.2 $\mu\text{g mL}^{-1}$; 13: SDMA 0.2 $\mu\text{g mL}^{-1}$; 14: SDMA- d_6 0.2 $\mu\text{g mL}^{-1}$; 15: glycine 9 $\mu\text{g mL}^{-1}$; 16: dimethylglycine 9 $\mu\text{g mL}^{-1}$; 17: cysteine 12 $\mu\text{g mL}^{-1}$; 18: homocysteine 3 $\mu\text{g mL}^{-1}$; 19: SAH 0.2 $\mu\text{g mL}^{-1}$; 20: SAM 0.2 $\mu\text{g mL}^{-1}$

getting cleaner plasma extracts and removing possible interferences: Oasis MCX 3 cc (60 mg) and Hybrid SPE

Phospholipid Ultra 30 mg/1 mL cartridges. However, the recovery was not good enough with neither of them, and for that

Table 2 Retention times, molecular formulas, molecular exact masses, precursor ions, and accurate m/z ratios in MS and MS/MS mode for each analyte examined

RT (min)	Compound	Molecular formula	MW	Precursor ion	m/z
2.40	Glycine	$\text{C}_2\text{H}_5\text{O}_2\text{N}$	75.0320	$[\text{M}+\text{H}]^+$	76.0393
2.43	Citrulline	$\text{C}_6\text{H}_{13}\text{O}_3\text{N}_3$	175.0957	$[\text{M}+\text{H}]^+$	176.1030
2.46	Dimethylglycine	$\text{C}_4\text{H}_9\text{O}_2\text{N}$	103.0633	$[\text{M}+\text{H}]^+$	104.0706
2.50	Betaine	$\text{C}_5\text{H}_{11}\text{O}_2\text{N}$	117.0790	$[\text{M}+\text{H}]^+$	118.0865
2.52	Cysteine	$\text{C}_3\text{H}_7\text{O}_2\text{NS}$	121.0197	$[\text{M}+\text{H}]^+$	122.0270
2.54	Creatine	$\text{C}_4\text{H}_9\text{O}_2\text{N}_3$	131.0695	$[\text{M}+\text{H}]^+$	132.0768
2.54	Creatine- d_3	$\text{C}_4\text{H}_6\text{D}_3\text{O}_2\text{N}_3$	134.0883	$[\text{M}+\text{H}]^+$	135.0956
2.59	Glutathione	$\text{C}_{10}\text{H}_{17}\text{O}_6\text{N}_3\text{S}$	307.0838	$[\text{M}+\text{H}]^+$	308.0911
2.59	Glutathione- $^{13}\text{C}_2^{15}\text{N}$	$\text{C}_8^{13}\text{C}_2\text{H}_{17}\text{O}_6\text{N}_2^{15}\text{NS}$	310.0876	$[\text{M}+\text{H}]^+$	311.0954
2.71	Homocysteine	$\text{C}_4\text{H}_9\text{O}_2\text{NS}$	135.0354	$[\text{M}+\text{H}]^+$	136.0427
3.08	Arginine	$\text{C}_6\text{H}_{14}\text{O}_2\text{N}_4$	174.1117	$[\text{M}+\text{H}]^+$	175.1193
3.11	Methionine	$\text{C}_5\text{H}_{11}\text{O}_2\text{NS}$	149.0510	$[\text{M}+\text{H}]^+$	150.0583
3.23	Creatinine	$\text{C}_4\text{H}_7\text{ON}_3$	113.0589	$[\text{M}+\text{H}]^+$	114.0667
3.23	Creatinine- d_3	$\text{C}_4\text{H}_4\text{D}_3\text{ON}_3$	116.0777	$[\text{M}+\text{H}]^+$	117.0850
3.53	Homoarginine	$\text{C}_7\text{H}_{16}\text{O}_2\text{N}_4$	188.1273	$[\text{M}+\text{H}]^+$	189.1346
4.58	ADMA	$\text{C}_8\text{H}_{18}\text{O}_2\text{N}_4$	202.1430	$[\text{M}+\text{H}]^+ > [\text{M}-\text{C}_6\text{H}_{10}\text{O}_2\text{N}_3]^+$	203.1503 > 46.0657
4.74	SDMA	$\text{C}_8\text{H}_{18}\text{O}_2\text{N}_4$	202.1430	$[\text{M}+\text{H}]^+ > [\text{M}-\text{CH}_4\text{N}]^+$	203.1503 > 172.1086
4.74	SDMA- d_6	$\text{C}_8\text{H}_{12}\text{D}_6\text{O}_2\text{N}_4$	208.1806	$[\text{M}+\text{H}]^+ > [\text{M}-\text{CHD}_3\text{N}]^+$	209.1879 > 175.1274
5.37	SAH	$\text{C}_{14}\text{H}_{20}\text{O}_4\text{N}_6\text{S}$	384.1216	$[\text{M}+\text{H}]^+$	385.1289
8.52	SAM	$\text{C}_{15}\text{H}_{22}\text{O}_4\text{N}_6\text{S}$	398.1372	$[\text{M}+\text{H}]^+$	399.1445

reason, solid-phase extraction was discarded, and protein precipitation was selected as the unique technique for extraction.

As previously mentioned, some of the compounds are amino thiols, and in order to determine their total concentration, the use of a reducing agent like tri-*n*-butylphosphine (TNBT) [32–34], sodium borohydride (NaBH₄) [35–39], tris(2-carboxyethyl)phosphine (TCEP) [30, 40], triphenylphosphine (TPP) [41], or 2-mercaptoethanol (2-ME) [42] is necessary. In this case, 2-mercaptoethanol (2-ME) and dithiothreitol (DTT) were assayed to reduce disulfide bonds and avoid their binding to proteins and other amino thiols. Concerning 2-mercaptoethanol, 5 μL of 2-ME 15 % was added to 50 μL of plasma and it was incubated for 30 min at 37 °C [42], and then, protein precipitation and the subsequent evaporation and reconstitution in ammonium formate 5 mM were carried out. On the other hand, the addition of 50 μL of dithiothreitol 77 g L^{-1} to 50 μL of plasma and the incubation of 15 min at room temperature were performed [26, 43]. As the preliminary studies did not show any relevant difference between both of them, dithiothreitol was chosen for being a more simple procedure, for avoiding heating the sample and because of the reduced odor and lower toxicity comparing with 2-mercaptoethanol [44].

The effect of DTT in amino thiols and in the rest of the analytes was assayed by comparing the signals obtained when the same plasma was treated with and without DTT. For glutathione and cysteine, the signal in untreated plasma was tiny in comparison with the signal obtained in the plasma aliquots treated with DTT. Moreover, in the case of cysteine, when using DTT, its dimer cystine, which ionize as the protonated form, disappeared totally proving its complete reduction. For homocysteine, the difference observed with DTT and without DTT was not especially considerable. Taking into account that around 80–90 % of the homocysteine is bound to proteins [45] and according to previous research [37], in this case, the breaking of homocysteine to protein bond when protein precipitation is carried out could be suggested. Besides, according to some studies, the use of a reduction agent provides a longer stability to the sample [37].

Regarding the rest of the compounds, homoarginine was the unique compound which undergoes significant loss of signal when using DTT. Nevertheless, the remaining area is more than enough to quantify homoarginine in plasma, as concentrations around a few $\mu\text{g mL}^{-1}$ were expected.

Taking into account that some analytes present in plasma are in concentrations above the linearity range, the selection of a dilution factor of the sample was essential. For that reason, 1:2, 1:4, and 1:10 dilution of plasma was assayed by evaporating and reconstituting the sample in higher volume of ammonium formate 5 mM. Finally, doing a dilution of 1:4 by taking 50 μL of plasma samples or calibration standards and reconstituting them in 200 μL of ammonium formate was decided, in order to achieve the best relation in decreasing

matrix effect and at the same time allowing the detection of SAH and SAM which were expected to be in the range of a few ng mL^{-1} in real samples.

The recovery of the extraction process was assessed by adding isotopically labeled compounds before and after the extraction process for two different concentration levels. For glutathione-¹³C₂¹⁵N, the recovery was on average 78.5 %, whereas for creatine-d₃ the 88.2 % of the analyte was recovered after the extraction process. Finally, for creatinine-d₃ and SDMA-d₆, the highest recoveries, 97.1 and 91.5 %, respectively, were obtained.

Method validation

Amino acids, amino acid derivatives, and related compounds are endogenous compounds, which appear in different concentrations in plasma. Due to the difficulty in obtaining blank matrices free of analytes, different approaches were carried out in order to select a proper quantification method.

First, the possibility of using stock solutions prepared in ammonium formate or in Serasub[®] serum surrogate was considered. In order to confirm whether the matrix effect was the same for ammonium formate, Serasub[®], and plasma, two different levels of isotopically labeled compounds were spiked in the same quantity in all the matrices, and the obtained areas were compared. Lower level matrices contained 1.5 $\mu\text{g mL}^{-1}$ of creatinine-d₃, creatine-d₃, and glutathione-¹³C₂¹⁵N and 50 ng mL^{-1} of SDMA-d₆, whereas upper level matrices contained 3 $\mu\text{g mL}^{-1}$ of creatinine-d₃, creatine-d₃, and glutathione-¹³C₂¹⁵N and 100 ng mL^{-1} of SDMA-d₆.

Matrix factor (MF) was calculated by dividing the signal of each spiked analyte in plasma versus the result of fortification with the same amount of analyte in ammonium formate solution. Similarly, MF was calculated for synthetic serum versus ammonium formate solution as well, resulting in different matrix factors for both matrices. The MF for isotopically labeled glutathione was 39.8 and 95.5 % for plasma and Serasub[®] respectively, for creatine-d₃ was 56.2 and 18.5 %, for creatinine-d₃ was 81.9 and 58.8 % and for SDMA-d₆ was 91.3 and 83 %. Huge differences were found in all cases except for SDMA-d₆, which has similar area for the three matrices in both concentration levels. Besides, when comparing plasma with Serasub[®], for all the compounds, the difference ranged from 2.4-fold higher area to 3-fold smaller area when comparing plasma with Serasub[®].

For all these reasons, the use of ammonium formate or Serasub[®] in terms of quantification was avoided and standard addition quantification for each sample was considered as the best alternative instead. Nevertheless, sample volume was not enough for carrying out all the additions necessary for a good quantification. Thus, a plasma pool made from both CKD samples and control samples, previously quantified by

standard addition, was used to prepare the calibration in plasma as detailed in *Standards and QC samples* section.

Linear regression equations were obtained by plotting the relation between the analytes and the internal standards versus analyte concentration. Correlation coefficients were in all cases above 0.9936. However, plasma concentrations below the first level of the calibration were expected in some samples. Therefore, it was necessary to prove that the linearity of the calibration in plasma pool was translatable to lower concentration levels. With this aim, the linearity of the method was assayed by using isotopically labeled compounds and the linearity was proved for a calibration range from 100 ng mL⁻¹ to 50 µg mL⁻¹ for glutathione-¹³C₂¹⁵N, from 5 ng mL⁻¹ to 20 µg mL⁻¹ for creatinine-d₃, from 5 ng mL⁻¹ to 50 µg mL⁻¹ for creatine-d₃, and from 5 ng mL⁻¹ to 50 µg mL⁻¹ for SDMA-d₆. The LOQs for creatine and glutathione were similar to those from recent works dealing with LC-MS in plasma matrix [46, 47], while the LOQs for creatinine and SDMA were 5 and 2-fold lower, respectively [48, 49].

Based on this, limits of quantification (LOQ) were estimated as follows: 100 ng mL⁻¹ for glutathione, 5 ng mL⁻¹ for creatine and creatinine in MS mode, and 5 ng mL⁻¹ for SDMA in MS/MS mode. For the rest of the compounds, isotopically labeled analogues were not available for the estimation of this parameter.

The accuracy and precision of the proposed method were calculated for unspiked, low, medium, and high QC samples. For that purpose, five determinations were carried out for each level and were repeated for 3 days. As shown in Table 3, accuracy fell within the acceptable range in this study and was between 83.2 and 116.3 % in all cases. Likewise, taking into account the closeness of repeated individual measures of each analyte, the RSD of the repeatability was below 13.7 % for all the analytes, whereas intermediate precision was below 16.7 % for all the compounds.

Regarding the stability, individual stock solutions were stable for more than a month. However, combined stock solutions were stable for at least 48 h at room temperature (86–101 %) and frozen for at least 1 week (82–114 %), but for less than a month (54–121 %). For that reason, combined stock solutions were prepared weekly from individual stock solutions, which proved to be stable for longer. Plasma extracts were proved to be stable at room temperature for 24 h (89–115 %), except for cysteine, which was stable for 12 h. In the same way, plasma extracts showed to be stable for at least 1 week frozen (86–122 %), but only stable for one freeze and thaw cycle.

Analysis of real samples

After the validation of the analytical method, the analysis of real samples was performed. A chromatogram of a plasma sample from a CKD patient is shown in Fig. 3. As previously

reported, these samples consisted of both plasma from pediatrics suffering from CKD and controls. It has to be highlighted that the first group was heterogeneous, consisting of patients in different stages of the disease and some of them had received RRT. Concentrations of all the different analytes measured by the developed method are shown in Table 4. Descriptive statistics are shown as median and interquartile range (IQR). With the aim of finding the significance of increased or decreased compounds in CKD patients, first of all, Kolmogorov-Smirnov test was done to find which variables were under a normal distribution, in order to use parametric or non-parametric test. Student's *t* test was used for normal variables, whereas *U* test in accordance to Mann and Whitney was applied to non-normal variables.

Despite the varying nature of pediatrics involved in this study, when taking into account suffering from CKD versus not suffering from it, significant differences were found between plasma concentrations of some analytes. Glycine, citrulline, creatinine, SDMA, and ADMA are shown to be significantly increased ($p \leq 0.001$ for all of them) in patients with CKD.

It has to be highlighted that plasma creatinine concentration in patients suffering from CKD differed consistently, and some renal patients did not have increased creatinine levels. In fact, some of the patients were in early-intermediate stages of the disease, whereas some others were suffering from advanced stages. Moreover, some of the patients were being treated by hemodialysis and some other had received kidney transplant.

For that reason, in order to find a relationship between high concentrations of creatinine and the increase or decrease of different compounds, patients with plasma concentrations higher than 12 µg mL⁻¹ were considered (a total of 21 patients) [50] and compared with control population. As a result, significant differences were found for dimethylglycine in addition to the previously reported analytes, being $p < 0.001$ for all the variables.

When taking into account all the samples, regardless of their creatinine levels, significant cross-correlations between different variables with respect to creatinine were found. The highest correlation was found for creatinine and SDMA ($r = 0.908$, $p < 0.01$), citrulline ($r = 0.839$, $p < 0.01$), and finally, SAM ($r = 0.773$, $p < 0.05$). Furthermore, a correlation study was carried out between analytes which have a correlation with creatinine. The correlation between SDMA and SAM is the highest, with a correlation coefficient of $r = 0.924$ ($p < 0.05$). Citrulline, which has been discovered to be significantly increased in pediatrics with CKD, despite their GFR and whether they receive or not RRT, correlates well especially with SDMA ($r = 0.837$, $p < 0.01$) and SAM ($r = 0.739$, $p < 0.01$). Correlation between creatinine and citrulline is shown in Fig. 4.

Table 3 Summary of the accuracy and precision results of the LC-QTOF-MS method for the analyzed quality control samples

Compound	Within-run accuracy (%nominal value)				Within-run precision (%RSD)				Between-run precision (%RSD)			
	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC	Unspiked QC	Low QC	Medium QC	High QC
Glycine	111.6	100.9	83.2	93.8	3.6	13.7	4.1	5.5	4.5	11.0	10.5	9.4
Citrulline	98.8	97.0	83.7	95.7	1.9	5.7	3.6	3.4	6.8	4.0	5.1	2.6
Dimethylglycine	123.7	101.5	88.1	98.3	2.0	7.1	3.9	3.5	16.2	8.8	9.0	6.8
Betaine	106.3	94.8	90.5	116.3	2.0	3.5	1.4	3.1	7.2	6.6	7.6	5.1
Cysteine	110.9	97.6	83.6	97.7	4.2	1.2	9.2	3.6	13.7	14.4	14.4	16.0
Creatine	93.8	93.7	99.0	100.5	1.7	1.2	1.3	1.7	7.6	12.9	10.1	9.2
Glutathione	103.2	106.4	102.0	108.4	6.1	10.0	7.1	4.7	1.4	3.8	7.9	1.5
Homocysteine	91.2	98.6	100.7	102.5	5.7	3.4	2.6	4.2	5.3	0.9	4.8	4.5
Arginine	111.8	99.8	89.6	85.2	1.3	3.6	1.4	4.8	6.1	3.5	5.6	5.0
Methionine	112.7	97.5	101.5	100.1	1.2	3.7	3.7	5.7	9.4	9.1	3.3	4.1
Creatinine	100.2	96.7	89.3	92.6	0.8	0.4	0.5	1.2	0.6	0.3	0.3	1.2
Homoarginine	95.1	96.7	92.0	96.4	1.0	5.5	1.9	6.0	10.4	14.5	12.3	14.7
ADMA	96.3	100.3	100.6	99.0	5.8	1.4	1.0	4.9	12.4	9.5	4.7	10.0
SDMA	101.8	110.2	101.8	98.8	7.7	11.0	10.1	9.3	9.5	6.4	8.6	10.1
SAH	102.3	94.0	97.3	97.0	8.6	4.7	1.3	2.7	14.4	10.1	14.9	13.0
SAM	95.4	101.3	99.7	109.7	4.9	1.6	2.0	2.2	8.4	13.5	16.7	13.7

Although SAH and SAM were not statistically significant when doing *U* test in accordance to Mann and Whitney ($p > 0.05$), when comparing median values and IQR, slight differences appeared to be between both populations. For that reason, SAM/SAH ratio was calculated for all the samples, with the aim of discovering if methylation power in pediatrics is affected when suffering from CKD, but no significant difference was found between both groups ($p > 0.05$). Thus, it could be suggested that a higher population would be needed in order to clarify if there is any significant difference between both populations as the variance in both populations differed.

As for dimethylglycine being increased in patients with high creatinine levels, the correlation for creatinine and

dimethylglycine was also studied. Nevertheless, correlation between both of them was not as high as for the rest of the analytes ($r = 0.493$, $p < 0.01$). The same situation occurred for ADMA, which had a correlation with creatinine of $r = 0.438$ ($p < 0.01$). Therefore, it could not be proved whether they are dependent variables with respect to creatinine.

Regarding the role of these compounds in metabolism, they are related to arginine-creatine pathway, to arginine methylation, and to urea cycle. Moreover, the increase of glycine and citrulline plasma concentration in pediatrics with CKD could suggest low activities of the enzymes arginine/glycine amidinotransferase (AGAT) [51]. Besides, SDMA, the inactive isomer of ADMA, is formed by methylation of arginine

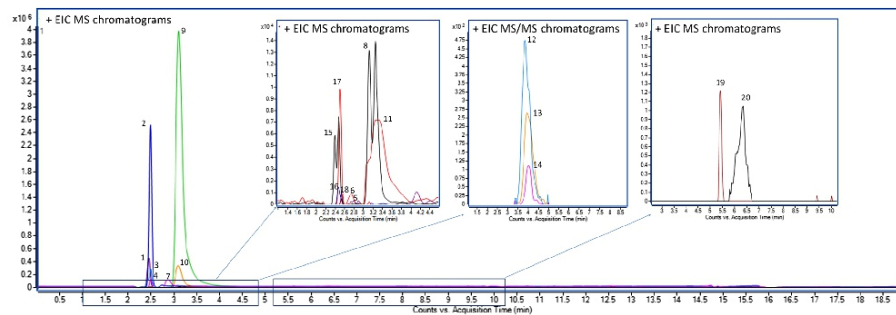
**Fig. 3** Example of a chromatogram of a CKD patient's plasma sample. Chromatographic peaks as numbered following the criteria in Fig. 2

Table 4 Results obtained for the analytes of interest in the developed method expressed as median (3rd–97th interquartile range)

Compound	Control	CKD
Glycine	2.4 (1.7–5.7)	4.3 (1.8–11.7)***
Citrulline	2.2 (1.2–4.0)	3.8 (2.2–9.7)***
Dimethylglycine	0.7 (0.2–1.9)	1.6 (0.4–4.0) ¹
Betaine	3.4 (1.1–5.6)	3.9 (0.8–8.8)
Cysteine	23.2 (3.8–39.4)	19.1 (6.0–46.8)
Creatine	10.0 (1.7–24.6)	9.6 (1.1–43.4)
Glutathione	1.9 (0.5–4.7)	1.9 (0.4–6.7)
Homocysteine	0.830 (0.075–1.5)	0.734 (0.056–2.6)
Arginine	14.7 (12.5–16.2)	13.3 (9.9–18.3)
Methionine	3.4 (1.4–8.2)	2.8 (1.4–5.8)
Creatinine	5.1 (2.5–8.1)	17.0 (6.6–63.4)***
Homoarginine	0.495 (0.192–0.777)	0.392 (0.104–1.7)
ADMA	0.133 (0.046–0.217)	0.180 (0.117–0.275)***
SDMA	0.075 (0.024–0.148)	0.350 (0.082–1.4)***
SAH	0.008 (0.001–0.035)	0.036 (0.006–0.319)
SAM	0.022 (0.003–0.043)	0.065 (0.040–0.157)

When considering only CKD patients with increased plasma creatinine concentrations ($>12 \text{ ng mL}^{-1}$), differences in dimethylglycine concentration were found ($p < 0.001$)

p value is expressed as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[52]. Despite the similarities between both compounds, the majority of ADMA is degraded by the enzyme dimethylarginine dimethylaminohydrolase (DDAH), being eliminated by renal excretion only the 20 %, whereas SDMA

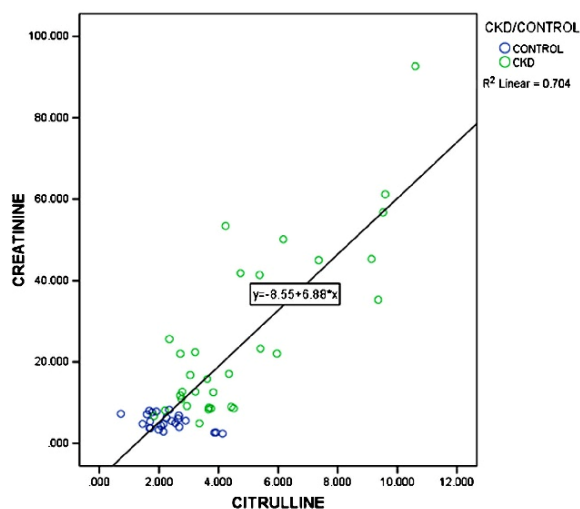
is mainly excreted in the urine [53]. Thus, SDMA accumulation could be suggested to occur due to the decreased excretion in renal patients. Concerning ADMA, it has been proposed that exaggerated endogenous synthesis of ADMA could occur because of its accumulation as well as a contributing factor to CKD progression [54]. Finally, the reason for the disturbed metabolism of dimethylglycine is still unknown.

Conclusions

A new LC-QTOF-MS method has been developed for the analysis of amino acids, amino acid derivatives, and related compounds taking part in the arginine-creatine pathway, the arginine methylation, and the urea cycle which could be of interest for the diagnosis of CKD in pediatric patients. This method enables the identification and quantification of 16 compounds in a single run in MS and MS/MS mode in less than 19 min. The sample treatment requires a low sample volume and involves the addition of dithiothreitol to obtain total aminothiols concentration, avoiding under- or overestimation in quantification, as well as a simple precipitation of plasma proteins without a previous derivatization step.

After optimization and validation, this method was successfully applied to the analysis of both a heterogeneous group of CKD pediatric patients and a control population, being significant differences found for glycine, citrulline, creatinine, ADMA, and SDMA, regardless of plasma creatinine concentrations, as well as for dimethylglycine in addition to the

Fig. 4 Correlation between creatinine and citrulline concentrations ($\mu\text{g mL}^{-1}$) in the analyzed CKD and control samples



previous compounds when increased creatinine levels were found in CKD pediatrics. Thus, it could be suggested that, apart from creatinine, which is currently used for the assessment of GFR, glycine, citrulline, creatinine, ADMA, and SDMA compounds could be potential biomarkers of the disease in pediatrics, regardless of their creatinine level or being or not treated with RRT, whereas these compounds as well as dimethylglycine could assess whether creatinine is increased, being potential indicatives of GFR as well. Future studies in a wider population could confirm whether these compounds are suitable biomarkers for the assessment of pediatric renal function in both the experimental and clinical setting.

The simple sample preparation step, the short analysis time, and the selectivity due to the mass accuracy of the LC-QTOF make this method suitable for the analysis of the whole arginine-creatine pathway, methylation cycle, and related compounds not only in chronic renal disease but also interesting for other pathologies such as cardiovascular disease or oxidative stress.

Acknowledgments The authors thank the technical and human support provided by Alava Central Service of Analysis belonging to SGiker (UPV/EHU, MINECO, GV/EJ, ERDF, and ESF) as well as the Division of Metabolism belonging to Cruces University Hospital (Barakaldo, Spain) for supplying real samples for this study. This work was funded by the Department of Industry, Innovation, Commerce and Tourism of the Basque Government (SAI12/25 Project) and by the Basque Government, Research Groups of the Basque University System (Project No. IT338-10). The Basque Government is also gratefully acknowledged for a predoctoral PRE_2013_I_899 grant for Sandra Benito (Department of Education, Language Policy and Culture).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Harambat J, van Stralen KJ, Kim JJ, Tizard EJ (2012) Epidemiology of chronic kidney disease in children. *Pediatr Nephrol* 27:363–373
- Kobayashi T, Yoshida T, Fujisawa T, Matsumura Y, Ozawa T, Yanai H, Iwasawa A, Kamachi T, Fujiwara K, Kohno M, Tanaka N (2014) A metabolomics-based approach for predicting stages of chronic kidney disease. *Biochem Biophys Res Commun* 445:412–416
- Avner ED, Harmon WE, Niaudet P, Yoshikawa N (2009) *Pediatric nephrology*, 6th edn. Springer, Berlin
- Tonshoff B, Schaefer F, Mehls O (1990) Disturbance of growth hormone-insulin-like growth factor axis in uraemia. Implications for recombinant human growth hormone treatment. *Pediatr Nephrol* 4:654–662
- Levey AS, Eckardt K-U, Tsukamoto Y, Levin A, Coresh J, Rossert J, De ZD, Hostetter TH, Lameire N, Eknoyan G (2005) Definition and classification of chronic kidney disease: a position statement from kidney disease: improving global outcomes (KDIGO). *Kidney Int* 67:2089–2100
- Noordzij M, Kramer A, Abad Diez JM, Alonso de la Torre R, Arcos Fuster E, Bikbov BT, Bonthuis M, Bouzas Caamaño E, Cala S, Caskey FJ, Castro de la Nuez P, Cemevskis H, Collart F, Diaz Tejero R, Djukanovic L, Ferrer-Alamar M, Finne P, Garcia Bazaga MdlA, Garneata L, Golan E, Gonzalez Fernández R, Heaf JG, Hoitsma A, Ioannidis GA, Kolesnyk M, Kramer R, Lasalle M, Leivestad T, Lopot F, van de Luijngaarden MWM, Macário F, Magaz Á, Martín Escobar E, de Meester J, Metcalfe W, Ots-Rosenberg M, Palsson R, Piñera C, Pippias M, Prütz KG, Ratkovic M, Resić H, Rodríguez Hernández A, Rutkowski B, Spustová V, Stel VS, Stojceva-Taneva O, Süleymanlar G, Wanner C, Jager KJ (2014) Renal replacement therapy in Europe: a summary of the 2011 ERA-EDTA Registry Annual Report. *Clin Kidney J*
- Chesnaye N, Bonthuis M, Schaefer F, Groothoff JW, Verrina E, Heaf JG, Jankauskiene A, Lukosiene V, Molchanova EA, Mota C, Peco-Antic A, Ratsch I-M, Bjerre A, Roussinov DL, Sukalo A, Topaloglu R, Van HK, Zagozdzon I, Jager KJ, Van SKJ (2014) Demographics of paediatric renal replacement therapy in Europe: a report of the ESPN/ERA-EDTA registry. *Pediatr Nephrol* 29:2403–2410
- Fanos V, Fanni C, Ottonello G, Noto A, Dessi A, Mussap M (2013) Metabolomics in adult and pediatric nephrology. *Molecules* 18: 4644–4857
- Wasung ME, Chawla LS, Madero M (2015) Biomarkers of renal function, which and when? *Clin Chim Acta* 438:350–357
- Bjornsson TD (1979) Use of serum creatinine concentrations to determine renal function. *Clin Pharmacokinet* 4:200–222
- Bosch JP, Saccaggi A, Lauer A, Ronco C, Belledonne M, Glabman S (1983) Renal functional reserve in humans. Effect of protein intake on glomerular filtration rate. *Am J Med* 75:943–950
- Waikar SS, Betensky RA, Bonventre JV (2009) Creatinine as the gold standard for kidney injury biomarker studies? *Nephrol Dial Transplant* 24:3263–3265
- Jain AK, McLeod L, Huo C, Cuerden MS, Akbari A, Tonelli M, van Walraven C, Quinn RR, Hemmelgam B, Oliver MJ, Li P, Garg AX (2009) When laboratories report estimated glomerular filtration rates in addition to serum creatinines, nephrology consults increase. *Kidney Int* 76:318–323
- Zhang M, Fang C, Smagin G (2014) Derivatization for the simultaneous LC/MS quantification of multiple neurotransmitters in extracellular fluid from rat brain microdialysis. *J Pharm Biomed Anal* 100:357–364
- Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens J-P, Morla A, Bouchu D (2003) ESI-MS/MS analysis of underivatized amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionization mode. *Rapid Commun Mass Spectrom* 17:1297–1311
- Uutela P, Ketola RA, Piepponen P, Kostiaainen R (2009) Comparison of different amino acid derivatives and analysis of rat brain microdialysates by liquid chromatography tandem mass spectrometry. *Anal Chim Acta* 633:223–231
- Mengerink Y, Kutlan D, Toth F, Csampai A, Molnar-Perl I (2002) Advances in the evaluation of the stability and characteristics of the amino acid and amine derivatives obtained with the o-phthalaldehyde/N-acetyl-L-cysteine reagents. High-performance liquid chromatography-mass spectrometry study. *J Chromatogr A* 949:99–124
- van Eijk HMH, Rooyackers DR, Soeters PB, Deutz NEP (1999) Determination of amino acid isotope enrichment using liquid chromatography-mass spectrometry. *Anal Biochem* 271:8–17
- Gartenmann K, Kochhar S (1999) Short-chain peptide analysis by high-performance liquid chromatography coupled to electrospray

- ionization mass spectrometer after derivatization with 9-fluorenylmethyl chloroformate. *J Agric Food Chem* 47:5068–5071
20. Piraud M, Vianey-Saban C, Petritis K, Elfakir C, Steghens J-P, Bouchu D (2005) Ion-pairing reversed-phase liquid chromatography/electrospray ionization mass spectrometric analysis of 76 underivatized amino acids of biological interest: a new tool for the diagnosis of inherited disorders of amino acid metabolism. *Rapid Commun Mass Spectrom* 19:1587–1602
 21. Qu J, Wang Y, Luo G, Wu Z, Yang C (2002) Validated quantitation of underivatized amino acids in human blood samples by volatile ion-pair reversed-phase liquid chromatography coupled to isotope dilution tandem mass spectrometry. *Anal Chem* 74:2034–2040
 22. Armstrong M, Jonscher K, Reisdorph NA (2007) Analysis of 25 underivatized amino acids in human plasma using ion-pairing reversed-phase liquid chromatography/time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 21:2717–2726
 23. Nolin TD, McMenamin ME, Himmelfarb J (2007) Simultaneous determination of total homocysteine, cysteine, cysteinylglycine, and glutathione in human plasma by high-performance liquid chromatography: application to studies of oxidative stress. *J Chromatogr B Anal Technol Biomed Life Sci* 852:554–561
 24. Guan X, Hoffman B, Dwivedi C, Mathees DP (2003) A simultaneous liquid chromatography/mass spectrometric assay of glutathione, cysteine, homocysteine and their disulfides in biological samples. *J Pharm Biomed Anal* 31:251–261
 25. Rossi R, Milzani A, Dalle-Donne I, Giustarini D, Lusini L, Colombo R, Di Simplicio P (2002) Blood glutathione disulfide: in vivo factor or in vitro artifact? *Clin Chem* 48:742–753
 26. Tang YB, Teng L, Sun F, Wang XL, Peng L, Cui YY, Hu JJ, Luan X, Zhu L, Chen HZ (2012) Determination of glycine in biofluid by hydrophilic interaction chromatography coupled with tandem mass spectrometry and its application to the quantification of glycine released by embryonal carcinoma stem cells. *J Chromatogr B Anal Technol Biomed Life Sci* 905:61–66
 27. Wu SE, Huskey WP, Borchardt RT, Schowen RL (1983) Chiral instability at sulfur of S-adenosylmethionine. *Biochemistry* 22:2828–2832
 28. van de Merbel NC (2008) Quantitative determination of endogenous compounds in biological samples using chromatographic techniques. *TrAC Trends Anal Chem* 27:924–933
 29. COMMISSION DECISION of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2012) *Off J Eur Communities*
 30. Le A, Ng A, Kwan T, Cusmano-Ozog K, Cowan TM (2014) A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS). *J Chromatogr B Anal Technol Biomed Life Sci* 944:166–174
 31. Piraud M, Vianey-Saban C, Bourdin C, Acquaviva-Bourdain C, Boyer S, Elfakir C, Bouchu D (2005) A new reversed-phase liquid chromatographic/tandem mass spectrometric method for analysis of underivatized amino acids: evaluation for the diagnosis and the management of inherited disorders of amino acid metabolism. *Rapid Commun Mass Spectrom* 19:3287–3297
 32. Bald E, Glowacki R, Drzewoski J (2001) Determination by liquid chromatography of free and total cysteine in human urine in the form of its S-quinolinium derivative. *J Chromatogr A* 913:319–329
 33. Klepacki J, Brunner N, Schmitz V, Klawitter J, Christians U, Klawitter J (2013) Development and validation of an LC-MS/MS assay for the quantification of the trans-methylation pathway intermediates S-adenosylmethionine and S-adenosylhomocysteine in human plasma. *Clin Chim Acta* 421:91–97
 34. Rogalewicz F, Hoppilliard Y, Ohanessian G (2000) Fragmentation mechanisms of α -amino acids protonated under electrospray ionization: a collisional activation and ab initio theoretical study. *Int J Mass Spectrom* 195(196):565–590
 35. Avila MA, Garcia-Trevijano ER, Lu SC, Corrales FJ, Mato JM (2004) Methylthioadenosine. *Int J Biochem Cell Biol* 36:2125–2130
 36. Kusmierek K, Glowacki R, Bald E (2006) Analysis of urine for cysteine, cysteinylglycine, and homocysteine by high-performance liquid chromatography. *Anal Bioanal Chem* 385:855–860
 37. Refsum H, Helland S, Ueland PM (1985) Radioenzymic determination of homocysteine in plasma and urine. *Clin Chem* 31:624–628
 38. Liang X-S, Zhao F-Q, Hao L-X (2013) Research on stability of synthetic folic acid. *Adv Mater Res* 781–784:1215–1218, 1215 pp
 39. Zhang W, Li P, Geng Q, Duan Y, Guo M, Cao Y (2014) Simultaneous determination of glutathione, cysteine, homocysteine, and cysteinylglycine in biological fluids by ion-pairing high-performance liquid chromatography coupled with precolumn derivatization. *J Agric Food Chem* 62:5845–5852
 40. Balion C, Kapur BM (2011) Folate: clinical utility of serum and red blood cell analysis. *Am Assoc Clin Chem* 37:8–10
 41. Desiderio C, Cavallaro RA, De Rossi A, D'Anselmi F, Fuso A, Scarpa S (2005) Evaluation of chemical and diastereoisomeric stability of S-adenosylmethionine in aqueous solution by capillary electrophoresis. *J Pharm Biomed Anal* 38:449–456
 42. Valerio A, Baldo G, Tessari P (2005) A rapid method to determine plasma homocysteine concentration and enrichment by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 19:561–567
 43. Hellmuth C, Koletzko B, Peissner W (2011) Aqueous normal phase chromatography improves quantification and qualification of homocysteine, cysteine and methionine by liquid chromatography-tandem mass spectrometry. *J Chromatogr B Anal Technol Biomed Life Sci* 879:83–89
 44. Ruseva S, Lozanov V, Markova P, Girchev R, Mitev V (2014) In vivo investigation of homocysteine metabolism to polyamines by high-resolution accurate mass spectrometry and stable isotope labeling. *Anal Biochem* 457:38–47
 45. Jacobsen DW (1998) Homocysteine and vitamins in cardiovascular disease. *Clin Chem* 44:1833–1843
 46. Squellerio I, Tremoli E, Cavalca V (2011) Quantification of arginine and its metabolites in human erythrocytes using liquid chromatography-tandem mass spectrometry. *Anal Biochem* 412:108–110
 47. Wang J-m, Chu Y, Li W, Wang X-y, Guo J-h, L-l Y, Ma X-h, Ma Y-l, Q-h Y, C-x L (2014) Simultaneous determination of creatine phosphate, creatine and 12 nucleotides in rat heart by LC-MS/MS. *J Chromatogr B Anal Technol Biomed Life Sci* 958:96–101
 48. Midttun O, Kvalheim G, Ueland PM (2013) High-throughput, low-volume, multianalyte quantification of plasma metabolites related to one-carbon metabolism using HPLC-MS/MS. *Anal Bioanal Chem* 405:2009–2017
 49. Servillo L, Giovane A, D'Onofrio N, Casale R, Cautela D, Castaldo D, Balestrieri ML (2013) Determination of homoarginine, arginine, NMMA, ADMA and SDMA in biological samples by HPLC-ESI-mass spectrometry. *Int J Mol Sci* 14:20131–20138, 20138 pp
 50. Vanholder R, De SR, Glorieux G, Argiles A, Baummeister U, Brunet P, Clark W, Cohen G, De DPP, Deppisch R, Descamps-Latscha B, Henle T, Jorres A, Lemke HD, Massy ZA, Passlick-Deetjen J, Rodriguez M, Stegmayr B, Stenvinkel P, Tetta C, Wanner C, Zidek W (2003) Review on uremic toxins: classification, concentration, and interindividual variability. *Kidney Int* 63:1934–1943
 51. Andrade F, Rodriguez-Soriano J, Prieto JA, Aguirre M, Ariceta G, Martin S, Sanjurjo P, Aldamiz-Echevarria L (2008)

- The arginine-creatine pathway is disturbed in children and adolescents with renal transplants. *Pediatr Res* 64:218–222
52. Sjostrom PA, Odland BG, Wolgast M (1988) Extensive tubular secretion and reabsorption of creatinine in humans. *Scand J Urol Nephrol* 22:129–131
53. Namnum P, Insogna K, Baggish D, Hayslett JP (1983) Evidence for bidirectional net movement of creatinine in the rat kidney. *Am J Physiol* 244:F719–F723
54. Mihout F, Shweke N, Bige N, Jouanneau C, Dussaule J-C, Ronco P, Chatziantoniou C, Boffa J-J (2011) Asymmetric dimethylarginine (ADMA) induces chronic kidney disease through a mechanism involving collagen and TGF- β 1 synthesis. *J Pathol* 223:37–45

ANNEX II:

Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data



Plasma biomarker discovery for early chronic kidney disease diagnosis based on chemometric approaches using LC-QTOF targeted metabolomics data

S. Benito^a, A. Sánchez-Ortega^b, N. Unceta^a, J.J. Jansen^c, G. Postma^c, F. Andrade^d, L. Aldámiz-Echevarria^d, L.M.C. Buydens^c, M.A. Goicolea^a, R.J. Barrio^a

^a Department of Analytical Chemistry, University of the Basque Country (UPV/EHU), Faculty of Pharmacy, Paseo de la Universidad 7, 01006 Vitoria-Gasteiz, Spain ^b Central Service of Analysis (SGiker), University of the Basque Country (UPV/EHU), Laskarai Ikerunea, Miguel de Unamuno 3, 01006 Vitoria-Gasteiz, Spain

^c Radboud University, Institute for Molecules and Materials (Analytical Chemistry-Chemometrics), P.O. Box 9010, 6500 GL Nijmegen, The Netherlands ^d Group of Metabolism, BioCruces Health Research Institute, CIBER de Enfermedades Raras (CIBERER), Plaza de Cruces 12, 48903 Barakaldo, Spain

ARTICLE INFO

Article history:

Received 26 April 2017
Received in revised form 10 October 2017
Accepted 28 October 2017 Available online 29 October 2017

Keywords:

Metabolomics
Biomarker
Chronic kidney disease
Citrulline
S-adenosylmethionine
Symmetric dimethylarginine

ABSTRACT

Chronic kidney disease (CKD) is a progressive pathological condition in which renal function deteriorates in time. The first diagnosis of CKD is often carried out in general care attention by general practitioners by means of serum creatinine (CNN) levels. However, it lacks sensitivity and thus, there is a need for new robust biomarkers to allow the detection of kidney damage particularly in early stages. Multivariate data analysis of plasma concentrations obtained from LC-QTOF targeted metabolomics method may reveal metabolites suspicious of being either up-regulated or down-regulated from urea cycle, arginine methylation and arginine-creatinine metabolic pathways in CKD pediatrics and controls. The results show that citrulline (CIT), symmetric dimethylarginine (SDMA) and S-adenosylmethionine (SAM) are interesting biomarkers to support diagnosis by CNN: early CKD samples and controls were classified with an increase in classification accuracy of 18% when using these 4 metabolites compared to CNN alone. These metabolites together allow classification of the samples into a definite stage of the disease with an accuracy of 74%, being the 90% of the misclassifications one level above or below the CKD stage set by the nephrologists. Finally, sex-related, age-related and treatment-related effects were studied, to evaluate whether changes in metabolite concentration could be attributable to these factors, and to correct them in case a new equation is developed with these potential biomarkers for the diagnosis and monitoring of pediatric CKD.

© 2017.

1. Introduction

Chronic kidney disease (CKD) is a major worldwide public health problem, affecting both children and adults, in which kidney function declines progressively. In clinical practice, different equations based on creatinine (CNN) concentration are used to estimate the glomerular filtration rate (GFR). This value reflects kidney function and is useful for the diagnosis of CKD and for assigning the stage or degree of the disease. Detection of CKD is considered a priority for primary care, because early treatment of CKD and its complications may delay or prevent the development of end-stage renal disease (ESRD) [1]. In clinical practice, one of the most commonly used equation in pediatrics to diagnose the disease for the first time is the Schwartz formula which is based on height, serum CNN concentration and *k* coefficient [2], as shown in Eq. (1).

$$GFR \text{ (mL min}^{-1}\text{/1.73 m}^2\text{)} = k \times \text{Height (cm)}/\text{Serum creatinine (mg dL}^{-1}\text{)} \quad (1)$$

Corresponding author.

Email address: r.barrio@ehu.es (R.J. Barrio)

<https://doi.org/10.1016/j.jpba.2017.10.036>
0731-7085/© 2017.

k coefficient has changed along the years, and is 0.45 for first year term infants, 0.55 for children and adolescent girls and 0.7 for adolescent boys currently.

Even if CNN is the classic biomarker used for the assessment of renal function in primary care attention, it has several drawbacks. Indeed, it lacks sensitivity and often reveals kidney damage when an important nephronic loss has already occurred. For that reason, in several early CKD patients a proper diagnosis by the general practitioner (GP) using the available CNN-based screening blood or urine tests is not possible until the disease progresses or more specific tests like abdominal computed tomography scan, abdominal ecography, kidney histopathology and immunohistochemistry or renal scintigraphy are carried out by nephrologists [3].

It would be ideal for GPs to be able to carry out better screening including more sensitive biomarkers for CKD in addition to CNN, as it is estimated that the majority of the population visits their GP within a 3-year period and can be subjected to screening [1]. Screening tests can be performed using either urine or blood biofluids. The utility of urinalysis is at times overestimated due to the inaccuracies in quantitatively collection of urine [4]. For that reason, blood analysis is preferred in children for diagnostic purposes. Thus, there is a need for new biomarkers (in addition to CNN) to be included in the equations used by GP in screening blood tests. Indeed, an earlier diagnosis of CKD, a better approximation to

of CKD, a better approximation to the CKD stage defined by nephrologists, monitoring of the progression of the disease and evaluation of the response to therapy are required. The early detection of CKD and the approximation of the CKD stage due to the implementation of new biomarkers in the screening tests carried out by GP would allow the early referral to the nephrologist, often leading to a better outcome of the patient.

CKD is associated with alterations in multiple metabolic pathways [5]. Arginine-creatinine metabolic pathway, arginine methylation and the urea cycle were suspicious to be affected in pediatric patients with CKD and thus, some metabolites from these metabolic pathways were expected to be increased or decreased in comparison to control pediatric patients. It has to be taken into account that depending on the metabolic pathways, differences in fold change concentration of metabolites can be lower or higher. For instance, the concentration of metabolites in the central metabolism is relatively constant. Concentrations of metabolites present in secondary metabolism-related pathways may differ more in concentration, depending on environmental conditions. Indeed, all biological systems are easily perturbed by a number of intra-individual or inter-individual experimental or environmental factors, such as age, diet, growth phase, media, nutrients, pH, sex, and temperature, which should be taken into consideration. This is known as induced biological variation [6]. Central metabolic pathways include glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, anaplerotic reactions and biosynthetic pathways of fatty acids and amino acids, and those reactions not included in central metabolic pathways are considered intermediate or secondary metabolic pathways [7]. The urea cycle is considered a central metabolic pathway, as it is part of arginine biosynthesis metabolic pathway, whereas arginine-creatinine metabolic pathway and arginine methylation are not included in the central metabolic pathway, and thus are considered secondary metabolic pathways.

Metabolomics aims at studying the dynamic changes, interactions and responses to stimuli of metabolites in different metabolic pathways [8]. The feasibility of metabolomics for biomarker discovery is supported by the assumption that metabolites play an important role in biological systems and that diseases cause disruption of biochemical pathways [9]. It has to be taken into account that each biofluid contains a large number of metabolites with concentration levels that can vary by orders of magnitude. However, from a biological point of view, metabolites present in high concentrations are not necessarily more important than those at low concentrations [6].

For all these reasons, in addition to the careful planning of experiments and analytical measurements, statistical and chemometric pre-processing are essential. Indeed, chemometrics, defined as the art of extracting chemically relevant information from data produced in chemical experiments, is indispensable to obtain consistent information and discard irrelevant information [10].

The aim of this work has been to perform data analysis on the plasma concentrations of 16 metabolites from arginine-creatinine, urea cycle and arginine methylation metabolic pathways in thirty-two patients at different stages of CKD and twenty-four control patients not suffering from CKD in order to find new potential biomarkers. These metabolites were selected because they were suspicious of being altered in pediatrics with CKD and were measured by means of a recently developed ion-pairing LC-QTOF-MS methodology targeted at these metabolites [11]. To turn these measurements into scores, we used complementary chemometric tools to extract the diagnostically relevant information that remain unseen with the naked eye. We have also identified the potential biomarkers for early diagnosis of CKD and checked whether these metabolites could be affected by age, sex or treatment received.

2. Material and methods

2.1. Chemicals and reagents

Acetonitrile used for both standard preparation and mobile phase was supplied by Scharlau (Sentmenat, Spain). In addition, LC-MS grade ammonium formate eluent additive from Fluka Analytical, Sigma-Aldrich (Steinheim, Germany) and perfluorooctanoic acid (PFHA) from Acros Organics (New Jersey, USA) completed the mobile phase. Standard preparation required the use of LC-MS grade methanol from Scharlau (Sentmenat, Spain), ultra-high purity water obtained from pretreated tap water by means of Elix reverse osmosis followed by a Milli-Q system from Millipore (Bedford, MA, USA) and chlorhydric acid, obtained from Merck (Darmstadt, Germany) as well.

Amino acid and amino acid derivative standards were supplied by different manufacturers. L-Cysteine (CYS), creatine (CTN), betaine (BET) and reduced glutathione (GSH) were supplied by TCI (Tokyo, Japan). L-Methionine (MET), L-arginine (ARG), glycine (GLY), L-homocysteine (HCYS), S-adenosyl-L-homocysteine (SAH), NG,NG'-dimethyl-L-arginine di(*p*-hydroxyazobenzene-*p*'-sulfonate) (SDMA), NG,NG'-dimethylarginine dihydrochloride (ADMA), S-adenosyl-L-methionine (SAM) and citrulline (CIT) were provided by Sigma-Aldrich (Steinheim, Germany). Fluorochem (Hadfield, UK) provided *N,N*-dimethylglycine (DMG) and homoarginine (HARG) and creatinine (CNN) was purchased from Alfa Aesar (Karlsruhe, Germany). Finally, creatine- d_3 H₂O from CDN Isotopes (Quebec, Canada) and SDMA- d_6 , and creatinine- d_3 (methyl- d_3) from Toronto Research Chemicals, TRC Canada (North York, Canada), and glutathione- $^{13}C_2^{15}N$ from Cambridge Isotope Laboratories (Andover, MA, USA) were used as internal standards. Dithiothreitol obtained from Fisher Scientific (Pittsburgh, PA, USA) was used as thiol reductant agent.

2.2. LC/QTOF method

Chromatographic analysis of plasma samples was carried out on an Agilent 1200 Series HPLC system coupled to Agilent 6530 Series hybrid quadrupole time-of-flight mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Agilent Jet Stream ESI was used as an ion source. Chromatographic separation and mass spectrometry are further described in a previous work [11]. This validated ion-pairing LC/QTOF methodology was used to quantify 16 metabolites from arginine-creatinine metabolic pathway, arginine methylation and urea cycle in these plasma samples: ADMA, ARG, BET, CIT, CYS, CTN, CNN, DMG, GSH, GLY, HARG, HCYS, MET, SAH, SAM and SDMA. It has to be noted that for the aminothiols (CYS, GSH and HCYS) total concentrations were quantified by means of a reduction process in sample treatment with dithiothreitol. Creatine- d_3 was used to correct the signal of GLY, CIT, DMG, BET, CYS and CTN, whereas glutathione- $^{13}C_2^{15}N$ corrected GSH signal. Similarly, creatinine- d_3 adjusted for the concentration of HCYS, ARG, MET, CNN and HARG, and SDMA- d_6 was used for ADMA, SDMA, SAM and SAH adjustment.

All the metabolites except for ADMA and SDMA were acquired in MS mode, with a scan rate of 2.2 spectra s^{-1} in 50–1000 m/z mass range. Taking into account that ADMA and SDMA are structural isomer compounds and elute at the same retention time, MS/MS mode with a scan rate of 3 spectra s^{-1} and a mass range from 30 to 1000 m/z was used with a collision energy of 10 V, which allows the detection of different ion products for each of the analytes. The protonated form $[M+H]^+$ of the analytes was detected for all the analytes

in MS mode, and the specific transitions 203.1503 > 46.0657 m/z and 203.1503 > 172.1086 m/z were monitored for ADMA and SDMA, respectively in MS/MS mode (Table 1). Fig. 1 shows an example chromatogram of a plasma sample containing both MS and MS/MS extracted ion chromatograms.

Regarding EU's decision 2002/657/CE proposed for the collection of identification points, confirmation of the presence of the analytes

Table 1
Retention times, accurate m/z ratios in MS and MS/MS modes and the internal standard (IS) used for signal correction for each examined analyte.

Compound	RT (min)	m/z	IS
GLY	2.40	76.0393	Creatinine-d ₃
CIT	2.43	176.1030	Creatinine-d ₃
DMG	2.46	104.0706	Creatinine-d ₃
BET	2.50	118.0865	Creatinine-d ₃
CYS	2.52	122.0270	Creatinine-d ₃
CTN	2.54	132.0768	Creatinine-d ₃
GSH	2.59	308.0911	Glutathione- ¹³ C ₂ ¹⁵ N
HCYS	2.71	136.0427	Creatinine-d ₃
ARG	3.08	175.1193	Creatinine-d ₃
MET	3.11	150.0583	Creatinine-d ₃
CNN	3.23	114.0667	Creatinine-d ₃
HARG	3.53	189.1346	Creatinine-d ₃
ADMA	4.58	203.1503 > 46.0657	SDMA-d ₆
SDMA	4.74	203.1503 > 172.1086	SDMA-d ₆
SAH	5.37	385.1289	SDMA-d ₆
SAM	8.52	399.1445	SDMA-d ₆
Glutathione- ¹³ C ₂ ¹⁵ N	2.59	311.0954	-
Creatinine-d ₃	2.54	135.0956	-
Creatinine-d ₃	3.23	117.0850	-
SDMA-d ₆	4.74	209.1879 > 175.1274	-

in samples is possible with this methodology, as retention time of each analyte and the accurate mass measurement of the m/z with an error less than 5 ppm are used [12].

2.3. Study samples

Thirty-two patients suffering from chronic kidney disease aged 3–17 years were recruited for this study according to the following criteria: subjects were followed up in Cruces University Hospital and clinically stable at the time of the study. The exclusion criteria include suffering from hepatopathy, anuria and/or insulin-dependent diabetes mellitus. The glomerular filtration rate (GFR) for each patient was calculated in the Pediatric Nephrology Service at Cruces University Hospital using Schwartz formula. Then, patients were classified by nephrologists according to glomerular filtration rate (GFR) and nephrological criteria based on complementary tests into stages CKD2 (60–89 mL/min/1.73 m²), CKD3 (30–59 mL/min/1.73 m²), CKD4 (15–29 mL/min/1.73 m²) and CKD5 (<15 mL/min/1.73 m²) of the disease [13]. In addition to the CKD stage, information regarding age, sex, and whether they received renal replacement therapy (dialysis or transplant) or not was available as well. For comparison purposes, 24 patients not suffering from chronic kidney disease aged 6–18 years were recruited for the study (Table 2).

Blood samples were withdrawn in the morning after an overnight fasting and were immediately cooled in an ice-water bath. Then, they were centrifuged at 1000g for 5 min at 4 °C. Samples were stored at -80 °C until sample treatment and analysis were carried out [11].

The study protocol was approved by Cruces Hospital Ethics Committee of Clinic Research and informed consent was given by patients' parents.

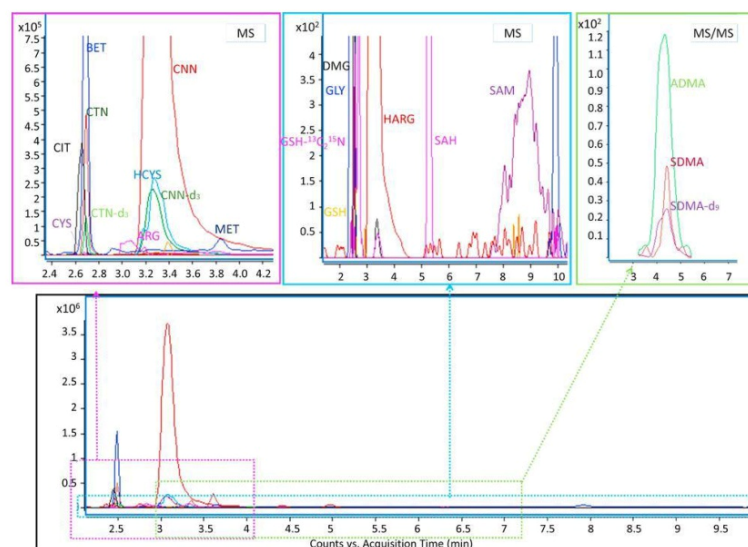


Fig. 1. Chromatogram of a plasma sample containing extracted ion chromatograms in MS/MS mode for ADMA and SDMA, and in MS mode for the rest of the analytes.

Table 2
Characteristics of the patients involved in the study.

CKD STAGE	Characteristics of the population			Number of patients (n)
	SEX	AGE	TREATMENT	
	(M/ F)	(2–12 y/ 13–18 y)	Not treated/Dialyzed/ Transplanted	
CONTROL	18/6	15/9	24/0/0	24
CKD2	8/6	10/4	9/0/5	14
CKD3	5/1	2/4	4/0/2	6
CKD4	2/4	3/3	5/0/1	6
CKD5	2/4	5/1	1/5/0	6

2.4. Sample treatment

First of all, samples were thawed and pooled plasma was made of both control and CKD samples to be used after quantifying it by standard additions to make a calibration curve in pooled plasma. In addition, 50 μL of plasma were placed in an Eppendorf tube to quantify each plasma samples in triplicate. Plasma samples were then spiked with 10 μL of a mix solution containing 15 $\mu\text{g mL}^{-1}$ of creatinine-d₃, creatine-d₃, and glutathione-¹³C₂¹⁵N and 5 $\mu\text{g mL}^{-1}$ of SDMA-d₆. Then, 50 μL of dithiothreitol (77,000 mg L^{-1}) were added and samples were incubated for 15 min at room temperature to perform the reduction of aminothiols compounds. In addition, 150 μL of cold acetonitrile were added to the mixture for plasma protein precipitation and was vortexed before centrifuging it for 10 min at 13,000 rpm at 4 °C. The obtained supernatant was then transferred to chromatographic vials, evaporated in nitrogen stream and the residue was reconstituted in 200 μL of ammonium formate 5 mM. Finally, these vials were transferred to autosampler for the analysis by means of the LC-QTOF method.

2.5. Chemometric analysis

2.5.1. Multivariate analysis

Principal Component Analysis (PCA) was carried out on logarithm transformed, autoscaled data obtained from the analysis of plasma samples with the developed LC-QTOF method. Matlab R2015a (Mathworks, Natick, Massachusetts, United States) software was used for performing the multivariate analysis. Outlying samples were de-fined according to Hierarchical Clustering with single, complete and average linkage and removed.

We compared several multivariate classification methods in their predictive ability for CKD, being Maximum Likelihood – Linear Discriminant Analysis (ML-LDA) [14], Fisher-Linear Discriminant Analysis (Fisher-LDA) [15], Quadratic Discriminant Analysis (QDA) [16] and *k* Nearest Neighbours (kNN) [17]. For that purpose, a random division of samples was done to obtain 80% of samples in a training set and 20% of the samples in a test set, and leave-one-out approach (LOO) was used on the training set to get the accuracy for each classification method. For kNN, it was necessary to select the optimum number of neighbours to be used, which was done by testing different *k* numbers of neighbours from 1 to 10 using LOO on the training set. The random division of the samples was repeated 50 times and the mean performances obtained for each classifier and each parameter using the leave-one-out approach on the training sets were used to compare the results.

Similarly, Partial Least Squares – Discriminant Analysis (PLS-DA) [18] was applied after optimizing the number of latent variables

(LV), and Sparse Partial Least Squares – Discriminant Analysis (sPLS-DA) [19] after selecting the number of metabolites to be included. In the case of sPLS-DA the number of latent variables was selected according to the number of groups minus one, as it is advised for generating the most stable models.

Once the classification method and the optimal parameters had been selected, they were applied to the samples from the test set and this operation was repeated 50 times to obtain the average performance of the selected model. To judge whether the 15 amino acids and related compounds measured improve the performance and add to the CNN-based diagnosis, the optimized performance of the classifiers were compared to the diagnostic performance by CNN alone.

3. Results and discussion

3.1. Summary statistics

To have an approximate idea of the orders of magnitude of the concentrations of the 16 metabolites analyzed in plasma, Fig. 2 shows an image made using Matlab R2015a (Mathworks Natick, Massachusetts, United States) representing in a color scale the concentration of each analyte in each patient's sample. In addition, a table summarizing concentration levels of each amino acid in plasma from both control and CKD patients is showed in the same figure.

This picture showed a big variation because even if almost all of the compounds were expected to be within the range from 0 to 100 μM , some metabolites like CNN had concentrations up to 800 μM . After re-scaling the colors in the color bar of the image from 0 to 100 μM , again 2 groups were differentiated: one containing analytes from 30 to 100 μM -and higher- (ARG, BET, CYS, CIT, CTN, CNN, DMG, GLY, GSH, HARG, HCYS, and MET) and another one containing metabolites approximately from 0 to 2 μM (ADMA, SDMA, SAH and SAM).

3.2. Multivariate analysis

3.2.1. Scaling and construction of PCA model

Multivariate data analysis tools take into consideration the intrinsic interdependency of the metabolite concentrations. As the concentrations of several metabolites were non-normally distributed across samples, logarithm transformation was carried out. Subsequently, autoscaling was performed by normalizing the concentration of each metabolite in each sample by subtracting mean metabolite concentration and dividing it by their standard deviation. This scaling method assumes that all metabolites are equally eligible to be important biomarkers, despite being present in higher or lower concentrations in plasma. One CKD sample was excluded for being an outlier according to Single Linkage, Complete Linkage and Average Linkage Hierarchical Clustering and was also visible as an outlier when doing Principal Component Analysis (PCA) plot.

This data pre-treatment enabled a separation between control and CKD sample groups when doing PCA, as shown in the biplot in Fig. 3.

Moreover, the relation between sample groups and metabolites can be obtained from the representation. SDMA, SAH, CNN, SAM, CIT, ADMA, GSH, DMG and GLY were found to be increased in CKD patients comparing with control samples. The first three principal components account for 58% of variance.

The information obtained from PCA was compared with the literature to check whether this up- or down-regulations could be explained by any known mechanism. SDMA and ADMA are thought to be increased due to the decreased excretion in renal patients. Moreover, the reduced activity of ADMA catabolism by dimethylarginine

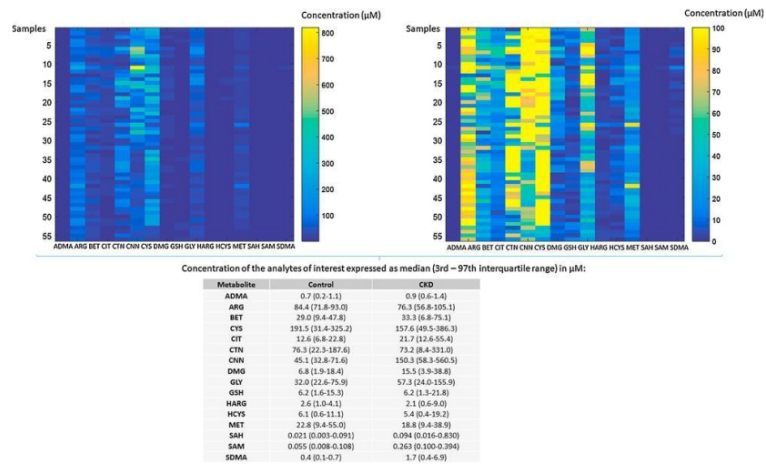


Fig. 2. Concentration of each amino acid per sample represented in a matrix according to a color scale (up) and a summary of the plasma concentrations for each metabolite per population group expressed as median and interquartile range in µM (down).

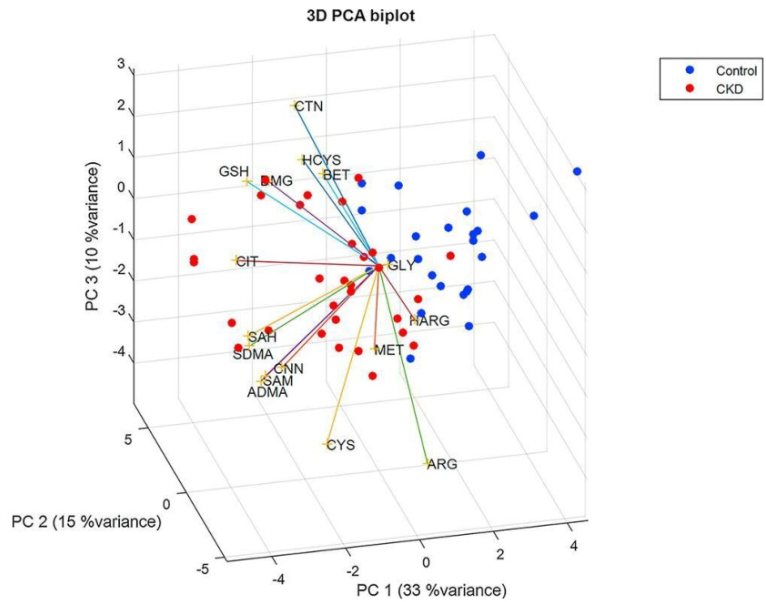


Fig. 3. PCA biplot containing the score plot for control and CKD samples and the loading plot with the 16 amino acids for the first 3 principal components.

dimethylaminohydrolyase (DDAH) would also be responsible for the increase of ADMA [20]. Regarding high CIT and GLY concentrations, low activity of the enzyme arginine/glycine amidinotransferase (AGAT) could be suggested [21]. Similarly SAH and SAM are thought to be increased in part due to impaired renal clearance, as kidneys play a major role in aminothiol metabolism [22]. Furthermore, it has been suggested that a reduction in the ability to metabolize and/or excrete SAH could be a primary event in CKD [23]. Several biomarkers of oxidative stress may be elevated in patients with CKD, such as glutathione [24], even if the nature of the oxidative stress in these patients still remains unclear. This impaired oxidative balance could be the result of increased reactive oxygen species (ROS) production and reduced clearance, or due to an ineffective antioxidant defense mechanism. Previous studies had already shown increased total GSH increased levels in adults suffering from CKD [25], but some other studies performed also in adults showed the opposite [26] or no difference between healthy and CKD patients [27]. Regarding DMG, it is a feedback inhibitor of betaine-homocysteine methyltransferase enzyme, which is normally excreted in urine or metabolized to sarcosine [28]. Thus, it is assumed that this increase could be related to accumulation, due to an impaired urinary excretion, in addition to any unknown factor that could also be involved.

The results of this unsupervised dimensionality reduction matched quite well with the results obtained using univariate analysis based on Student's *t* test for analytes with normal distribution and *U* test in accordance to Mann and Whitney for those with non-normal distribution after verifying the normal distribution of samples using Kolmogorov-Smirnov test. Indeed, the mean values of GLY, CIT, CNN, ADMA and SDMA were significantly increased in CKD patients when comparing with control individuals [11].

3.2.2. Classifying between early CKD and control samples using all metabolites

In order to find out whether any of these metabolites are interesting as potential biomarkers for the early diagnosis of CKD, a PCA model was constructed and a classification method was selected considering only early disease samples (those at CKD2 stage) and control samples (Fig. 4). Following these approaches, not only the separation between groups was complete except for one CKD2 sample which was placed with control samples, but also an improvement in the performance of the classification was gained when using all the metabolites including CNN, in comparison with the single use of CNN. Indeed, after the optimization of the classifiers, using all the metabolites with PLS-DA with 2 LV the percentage of success was 76% and with sPLS-DA with 1 LV and 4 metabolites was 67%. CIT,

CNN, SAM, SDMA were the output metabolites for sPLS-DA variable reduction and classification method. On the other hand, QDA was selected as the best classification method when using only CNN with a performance of 71%.

Moreover, in order to have an idea of how well all non-CNN metabolites performed in comparison with CNN, classification was also executed using the 15 metabolites, and a performance of 59% was obtained using PLS-DA with 2 LV. Therefore, even if CNN alone is better than the rest of the metabolites for prediction purposes, the rest of the metabolites could be useful biomarkers in combination with CNN.

As a consequence, CIT, SAM and SDMA and CNN were found to be the biologically relevant features according to the metabolite selection carried out with sPLS-DA, and it would match with the fact that CNN would be of special interest for classification purposes in comparison with the rest of the metabolites. Univariate statistical analysis was performed to verify if the two groups of samples were significantly different from each other for each of these 4 metabolites. The univariate statistical analysis based on the use of Student's *t*-test for normal distributed metabolites and *U* test in accordance to Mann-Whitney for non-normal ones showed *p*-values below 0.001 for CIT, SDMA and CNN. Despite the fact that SAM did not have a normal distribution and was not significant according to the tests ($p > 0.05$), median values and interquartile range showed slight differences between both populations, in addition to a correlation of SAM with SDMA ($r = 0.92$; $p < 0.05$), CIT ($r = 0.74$, $p < 0.01$) and CNN ($r = 0.77$, $p < 0.05$). Thus, the use of a higher population is suggested in order to clarify if there is any significant difference between both populations. Taking into account all of these evidences, it is worth using only these four metabolites as an input to check the accuracy of the different classifiers, described in the following sections.

3.2.3. Classifying between early CKD and control samples using CIT, SAM, SDMA and CNN

The most practical approach and the one of special interest would be to select the minimum number of metabolites capable of increasing prediction accuracy compared to CNN alone, and if possible also an increased one in comparison with using all the 16 metabolites. CNN, CIT, SAM and SDMA metabolites showed a similar efficiency for the early diagnosis of CKD in the output obtained from sPLS-DA in comparison with the prediction of CNN alone, as shown in the previous subsection. Furthermore, separation in the PCA model of CKD and control groups was almost complete (Fig. 5) and whether these metabolites used as an input could improve the accuracy of classification was checked as well. Once again, classifiers were optimized for

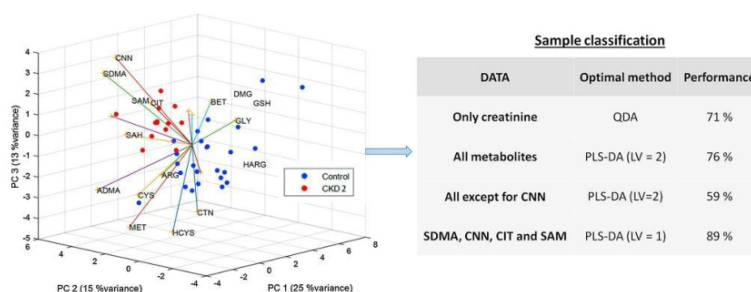


Fig. 4. PCA biplot for control and early CKD samples (on the left) and the performances for different sample classification methods (on the right).

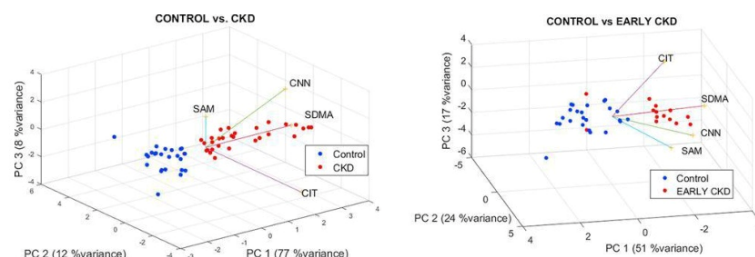


Fig. 5. PCA for CNN, CIT, SAH and SDMA data for control and CKD samples regardless of the stage of the disease (left) and for control and early CKD samples (right).

these 4 metabolites and a performance of 89% was obtained using PLS-DA (LV = 1) for the classification of early CKD and control samples. This means an increase in performance of 18% when using CIT, SAM and SDMA in addition to CNN in comparison with using CNN only.

3.2.4. Stage-independent CKD diagnosis using all metabolites

Although the main objective of this work was to find a few potential biomarkers for early diagnosis, it is also interesting to verify if these 16 metabolites could also be useful regardless of the stage of CKD to be applicable in clinical practice. For that reason, the same process was repeated for all the CKD samples, regardless of the CKD stage using the 16 metabolites.

First of all, the various classification methods were compared and PLS-DA (LV = 1) with a classification accuracy of 84% resulted in the best outcome using all the metabolites including CNN, whereas QDA was again the best classification method using CNN alone to sort the samples as control or CKD (81% performance). In both cases performance was similar, so using all metabolites for classification would neither improve the accuracy in the diagnosis nor jeopardize the performance.

3.2.5. Stage-independent CKD diagnosis using metabolites CIT, SAM, SDMA and CNN

As CIT, SAM, SDMA and CNN provide the best results for the early diagnosis of CKD, the classification procedure was repeated using as input the concentration information for these metabolites only in all the samples, regardless of the CKD stage. In this case, PLS-DA (LV = 1) proved to be the best classifier to be used and once applied to the total amount of samples a performance of 91% was achieved, thus enabling an improvement of 10% in comparison with CNN alone.

Accordingly, metabolites CIT, SAM and SDMA might be of special interest to add to CNN, as they improve the accuracy of CNN diagnosis not only for early disease samples but also when using all the samples, regardless of the CKD stage. Moreover, SDMA ($r = 0.91$, $p < 0.001$), citrulline ($r = 0.84$, $p < 0.001$) and SAM ($r = 0.77$, $p < 0.05$) were found to be correlated with CNN.

3.2.6. Classification performance for CKD stages

Within the years different equations have been developed to obtain the GFR with the aim of diagnosing those patients suffering from CKD and to classify them according to the stage of the disease, such as: Schwartz, Bedside Schwartz equation, Modification of Diet in Renal Disease (MDRD), Counahan-Barratt and Cockcroft-Gault [4]. For the development of new equations "gold standards" like inuline, io-

hexol, iohalamate, $^{51}\text{Cr-EDTA}$ or $^{99\text{m}}\text{Tc-DTPA}$ are commonly used to obtain a real GFR value and to adjust the equations to get a similar estimated GFR value based on the measurement of endogenous metabolites. Despite being this the ideal, as a preliminary study prior to the development of any new equation, metabolites were evaluated using PCA and different classification methods were used to determine the stage of the disease.

Concerning the differences in metabolite profile for different CKD stages, PCA model obtained from using all the metabolites showed that concentrations of some metabolites had increased with disease severity, such as SAH, SAM, SDMA, ADMA, CNN, CIT, GSH and DMG, as shown in Fig. 6. Moreover, there was a complete separation between the groups made of CTRL-CKD2 and CKD3 to CKD5 groups. Likewise, including only CIT, SAM, SDMA and CNN a good separation was achieved and even in the case in which CNN was not included, a gradation on CKD was observed (data not shown). To check whether all the metabolites or some of them could add, the accuracy of classification using the plasma concentrations of the proposed metabolites was compared with the use of CNN concentration only. Some increase in performance was found using all the 16 metabolites including CNN with PLS-DA and 2 LV (66%) and even a higher increase using CIT, SAM, SDMA and CNN with PLS-DA and 3 LVs (74%), in comparison with using only CNN with QDA classification

(52%). Even if in general these accuracies do not appear high, closer inspection of the confusion matrix shows that almost all misclassifications occur one stage above or below in comparison with the CKD stage assessed by the nephrologists and less than 3% of the samples are misclassified 2 stages or more above or below (Fig. 6). From a clinical point of view, misclassification of the samples into a different stage could affect the action plan chosen for the patient. However, these recommended approaches for each stage are general and do not change substantially when classifying the samples one stage above or below.

It has to be taken into account that in our predictions classification performance using plasma CNN alone might be that low because age, gender, height and/or weight of pediatrics are commonly used in different equations to correct for some factors affecting serum CNN concentration for the GFR calculation and there is a chance that the classification into the different CKD stages could be improved by using these factors, instead of using CNN as a linear predictor (Eq. (1)). For instance, Schwartz equation uses height to correct for CNN production, as it is a function of muscle mass which is related with body mass and among the variables of body size tested by them, body length di-vided by CNN concentration provided the best correlation with GFR [29].

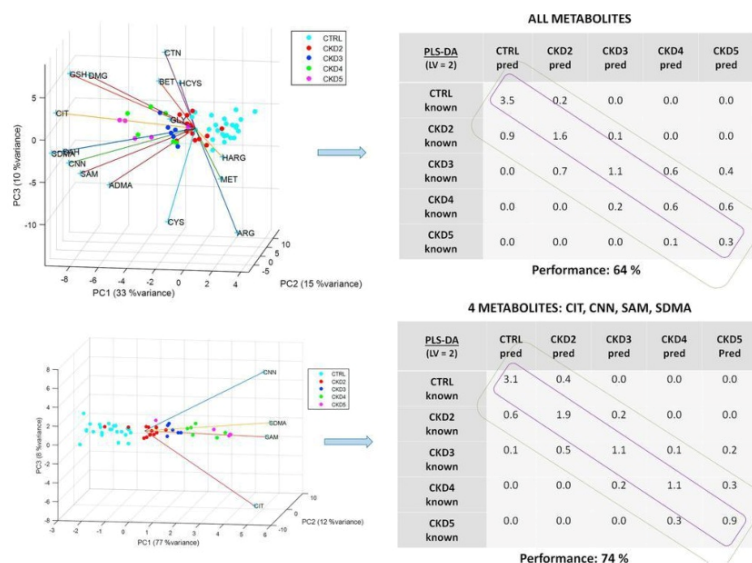


Fig. 6. PCA for all the metabolites showing the gradation of the samples according to the metabolite concentration and the stage of the CKD. In addition, the average confusion matrix for the best classifier for all the metabolites in 50 different test sets is shown (up). Similarly PCA for the 4 selected metabolites and its corresponding confusion matrix (down) are shown.

3.2.7. Age-, sex- and treatment-related effects

As mentioned before, it is necessary to check that the concentrations of these metabolites are not sex-related, age-related or treatment-related, in case these effects are present to be able to correct them in future equations, as it is done in Schwartz formula (Eq. (1)). For this purpose, PCA using CIT, CNN, SAM and SDMA was applied on the samples using two age groups ranging from 2 to 12 years old and from 13 to 18 years old, in agreement with the division made by Way et al., according to the normal GFR values shown for children [30]. Although there was no clear separation of the groups, some trend was observed, as shown in Fig. 7. Therefore, in order to find any relation between age and metabolite concentrations Partial Least Squares (PLS) regression method was carried out to predict the age according to the metabolite levels in plasma. No good prediction was obtained when all samples or only CKD samples were considered, whereas when only control samples were considered the prediction was much better, thus meaning that there is a positive correlation between age and metabolite concentrations in healthy pediatric. The results were similar for either using all the metabolites or using only CNN, SDMA, SAM and CIT as variables.

Regarding sex, samples from both male and female groups were overlapped in the score plot of the PCA, thus showing that there is not any difference in metabolite concentrations according to sex. This is something that could be expected since the patients were pediatric and no major differences in sex are presumed at these ages. In addition, the same operation was performed using only adolescent samples, that is, those aged between 13 and 18 years old. Nevertheless,

no clear separation was observed neither using all the metabolites nor the 4 selected ones.

Besides, to check if there could be some effect on metabolite concentration according to the treatment, PCA was carried out for untreated, transplanted and dialyzed CKD patients, but no separation was found between groups. However, transplanted patients showed a more homogeneous amino acid profile in comparison with not treated and dialyzed ones (Fig. 8). Therefore, it could be concluded that after receiving a transplant the amino acid profile tends to normalize and equalize. Moreover, if transplanted CKD patients are compared to control patients, the distribution of the samples is quite similar, thus implying that amino acid profile tends to equalize, even if both groups are still separated in PCA. This means that amino acid profiles from controls and transplanted patients have a similar distribution, but there are still some differences between both amino acid profiles.

This standardization in the metabolite concentration after transplant could be expected as patients were transplanted when other renal replacement therapy had failed and in all cases samples were collected at least a year after having received the transplant.

4. Conclusions

Three new metabolites, CIT, SAM and SDMA, have been proposed as potential biomarkers in addition to the commonly used CNN to be implemented since they enable a better diagnosis of early stages of CKD in pediatric. Moreover, these metabolites showed greater improvement also for the diagnosis of the rest of the stages in CKD. In addition, some gradation in the concentration of these metabolites according to the CKD stage has also been found. Therefore, it is rea-

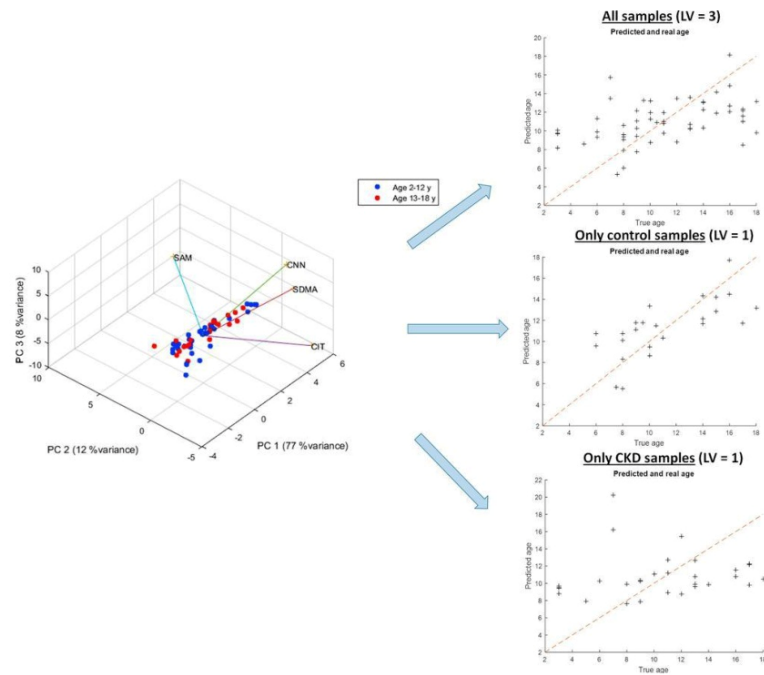


Fig. 7. On the left, PCA for all the samples with CIT, CNN, SAM and SDMA data, colored according to different age groups. On the right, a representation of a PLS regression to predict the age with all the samples, only control samples and only CKD samples.

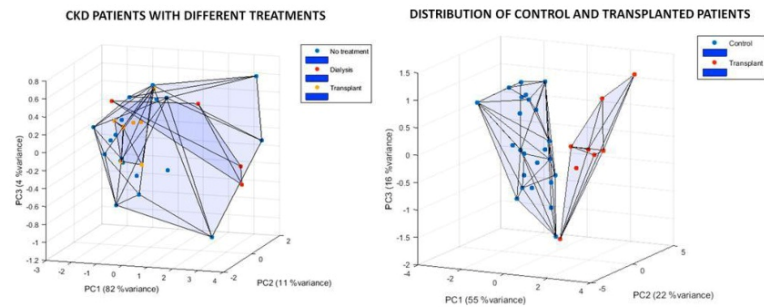


Fig. 8. On the left, PCA with convex hull showing different distributions of samples according to the concentrations of CIT, CNN, SAM and SDMA and the relation with the treatment received by CKD patient. On the right, sample distributions for control and transplanted patients.

sonable to think of including them in a new equation to be used in general practitioners' blood screening tests with some other commonly used parameters as cystatin C protein concentration, CNN concentration, age, height and weight, in order to be able to approximate the prediction of the disease performed by nephrologists and to

foresee the stage in CKD patients. In order to comply with this purpose, it is suggested to collect a higher number of samples, to obtain GFR by using some contrast agent like iohexol, which is used intravenously for radiologic procedures even in the presence of renal disease, and which is not secreted, metabolized or reabsorbed by the

kidney for future research. This would allow adjusting a new equation for GFR calculation including CIT, SDMA and SAM and comparing the performances obtained from this equation with other commonly used equations like Schwartz in pediatrics.

Disclosures

No relevant conflicts of interest to declare.

Acknowledgements

The authors thank for technical and human support provided by SGIker of UPV/EHU and European funding (ERDF and ESF) as well as the Division of Metabolism belonging to Cruces University Hospital (Barakaldo, Spain) for supplying real samples for this study. This work was funded by the Department of Industry, Innovation, Commerce and Tourism of the Basque Government (SAI 12/25 Project) and by the Basque Government, Research Groups of the Basque University System (Project No. IT3338-10). The Basque Government is also gratefully acknowledged for a predoctoral grant (PRE_2013_1_899) and for a mobility grant (EP_2016_1_0003) for Sandra Benito (Department of Education, Language Policy and Culture). This grant made possible doing an stay in Chemometrics group in the Analytical Chemistry Department at Radboud University, which is also thanked for the opportunity given. Dr. Ewa Szymanska is also gratefully acknowledged for kindly providing sPLSDA routine.

References

- [1] M. Rosenberg, R. Kalda, V. Kasilevicius, M. Lember, Management of chronic kidney disease in primary health care: position paper of the European Forum for Primary Care, *Qual. Prim. Care* 16 (2008) 279–294.
- [2] G.J. Schwartz, D.F. Work, Measurement and estimation of GFR in children and adolescents, *Clin. J. Am. Soc. Nephrol.* 4 (2009) 1832–1843.
- [3] S.F. Simoneaux, L.A. Greenbaum, Diagnostic imaging, in: E.D. Avner, W.E. Harmon, P. Niaudet, N. Yoshikawa (Eds.), *Pediatric Nephrology*, Springer, Berlin, 2009, pp. 535–564.
- [4] G.J. Schwartz, A. Munoz, M.F. Schneider, R.H. Mak, F. Kaskel, B.A. Warady, S.L. Furth, New equations to estimate GFR in children with CKD, *J. Am. Soc. Nephrol.* 20 (2009) 629–637.
- [5] A.D. Slee, Exploring metabolic dysfunction in chronic kidney disease, *Nutr. Metab.* 9 (2012) 36.
- [6] R.A. Van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, *BMC Genom.* 7 (2006).
- [7] C. Yang, A.D. Richardson, A. Osterman, J.W. Smith, Profiling of central metabolism in human cancer cells by two-dimensional NMR, GC-MS analysis, and isotopomer modeling, *Metabolomics* 4 (2008) 13–29.
- [8] A. Kalivodova, K. Hron, P. Filzmoser, L. Najdekr, H. Janeczkova, T. Adam, PLS-DA for compositional data with application to metabolomics, *J. Chemom.* 29 (2015) 21–28.
- [9] M.S. Monteiro, M. Carvalho, M.L. Bastos, P. Guedes de Pinho, Metabolomics analysis for biomarker discovery: advances and challenges, *Curr. Med. Chem.* 20 (2013) 257–271.
- [10] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics. A review in human disease diagnosis, *Anal. Chim. Acta* 659 (2010) 23–33.
- [11] S. Benito, A. Sanchez, N. Unceta, F. Andrade, L. Aldamiz-Echevarria, M.A. Goicolea, R.J. Barrio, LC-QTOF-MS-based targeted metabolomics of arginine-creatinine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease, *Anal. Bioanal. Chem.* 408 (2016) 747–760.
- [12] 2002/657/EC, Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, *Official Journal of the European Communities* (2002).
- [13] National Kidney Foundation, K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification, *Am. J. Kidney Dis.* 39 (2002) S1–S266.
- [14] T. Hastie, R. Tibshirani, J. Friedman, Linear methods for classification, in: T. Hastie, R. Tibshirani, J. Friedman (Eds.), *The Elements of Statistical Learning*, Springer, New York, 2008, pp. 101–118.
- [15] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugen.* 7 (2) (1936) 179–188.
- [16] S. Geisser, Posterior odds for multivariate normal distributions, *J. R. Stat. Soc. B Met.* 26 (1964) 59–76.
- [17] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, *IEEE T. Inf. Theory* IT 13 (1) (1967) 21–27.
- [18] S. Wold, M. Sjostrom, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst. Syst.* 58 (2001) 109–130.
- [19] E. Szymanska, E. Brodrick, M. Williams, A.N. Davies, H.-J. van Manen, L.M.C. Buydens, Data size reduction strategy for the classification of breath and air samples using multicapillary column-ion mobility spectrometry, *Anal. Chem.* 87 (2015) 869–875.
- [20] F. Mihout, N. Shweke, N. Bige, C. Jouanneau, J.-C. Dussaule, P. Ronco, C. Chatziantoniou, J.-J. Boffa, Asymmetric dimethylarginine (ADMA) induces chronic kidney disease through a mechanism involving collagen and TGF- β 1 synthesis, *J. Pathol.* 223 (2011) 37–45.
- [21] F. Andrade, J. Rodriguez-Soriano, J.A. Prieto, J. Elorz, M. Aguirre, G. Ariceta, S. Martin, P. Sanjujo, L. Aldamiz-Echevarria, The arginine-creatinine pathway is disturbed in children and adolescents with renal transplants, *Pediatr. Res.* 64 (2008) 218–222.
- [22] A. Valli, J.J. Carreno, A.R. Qureshi, G. Garibotto, P. Barany, J. Axelsson, B. Lindholm, P. Stenvinkel, B. Anderstam, M.E. Suliman, Elevated serum levels of S-adenosylhomocysteine, but not homocysteine, are associated with cardiovascular disease in stage 5 chronic kidney disease patients, *Clin. Chim. Acta* 395 (2008) 106–110.
- [23] K. Jabs, M.J. Koury, W.D. Dupont, C. Wagner, Relationship between plasma S-adenosylhomocysteine concentration and glomerular filtration rate in children, *Metab. Clin. Exp.* 55 (2006) 252–257.
- [24] M. Romeu, R. Nogueas, L. Marcus, V. Sanchez-Martos, M. Mulero, A. Martinez-Ven, J. Mulla, M. Giralt, Evaluation of oxidative stress biomarkers in patients with chronic renal failure: a case control study, *BMC Res. Notes* 3 (2010).
- [25] A. Zinella, S. Sotgia, G. Loriga, L. Deiana, A.E. Satta, C. Carru, Oxidative stress improvement is associated with increased levels of taurine in CKD patients undergoing lipid-lowering therapy, *Amino Acids* 43 (2012) 1499–1507.
- [26] I. Ceballos-Picot, V. Witko-Sarsat, M. Merad-Boudia, A.T. Nguyen, M. Thevenin, M.C. Jaudon, J. Zingraff, C. Verger, P. Jungers, B. Descamps-Latscha, Glutathione antioxidant system as a marker of oxidative stress in chronic renal failure, *Free Radic. Biol. Med.* 21 (1996) 845–853.
- [27] J. Himmelefarb, E. McMenamin, E. McMonagle, Plasma aminothiols oxidation in chronic hemodialysis patients, *Kidney Int.* 61 (2002) 705–716.
- [28] D.O. McGregor, W.J. Dellow, M. Lever, P.M. George, R.A. Robson, S.T. Chambers, Dimethylglycine accumulates in uremia and predicts elevated plasma homocysteine concentrations, *Kidney Int.* 59 (2001) 2267–2272.
- [29] G.J. Schwartz, G.B. Haycock, C.M. Edelmann Jr., A. Spitzer, A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine, *Pediatrics* 58 (1976) 259–263.
- [30] A.F. Way, A.M. Bolonger, J.G. Gambertogli, Pharmacokinetics and drug dosing in children with decreased renal function, in: E.D. Avner, M.A. Holliday, T.M. Barratt (Eds.), *Pediatric Nephrology*, Williams&Williams, Berlin, 1994, p. 1306.