Universitat Jaume I
Departament de Química Física i Analítica
Àrea de Química Analítica

Institut Universitari de Plaguicides i Aigües

POTENTIAL OF LC-(Q)TOF MS IN TARGET AND NON-TARGET ANALYSES: WIDE SCOPE SCREENING OF ORGANIC CONTAMINANTS AND METABOLOMIC APPLICATIONS

Tesis Doctoral
RAMON DÍAZ SAN PEDRO
2016

Dr. Félix Hernández Hernández y **Dr. Juan Vicente Sancho LLopís**, catedráticos de Química Analítica de la Universitat Jaume I de Castellón,

CERTIFICAN: que la Tesis Doctoral "Potential of LC-(Q)TOF MS in target and non-target analyses: wide scope screening of organic contaminants and metabolomic applications" ha sido desarrollada bajo su dirección, en el Instituto Universitario de Plaguicidas y Aguas, Departamento de Química Física y Analítica de la Universitat Jaume I de Castellón, por **Ramon Díaz San Pedro**.

Lo que certifico para los efectos oportunos en Castellón de la Plana, a 27 de Octubre de 2015.

Fdo. Dr. Félix Hernández Hernández Fdo.   Dr.Juan   Vicente   Sancho   LLopis

AGRADECIMIENTOS/ACKNOWLEDGMENTS

mi sobrino del alma, que me ha alegrado en tantas veces solo con una sonrisa y su admiración.

Por ultimo pero no por ello menos importante, Marian, la persona que más allá de estar en este apartado debería ser co-autora de este trabajo. Ella ha mantenido la llama viva cuando los vientos de Holanda asomaban en apagarla. Gracias morchi por todo la ayuda, por soportarme en los malos momentos, por entenderme y hacerme verlo todo de modo optimista, por nunca rendirte. Mil gracias por hacer que este sueño se haga realidad.

En definitiva, gracias a todos las personas que he ido encontrando en este largo camino, ya que, de una y otra forma, han contribuido en esta Tesis.

RESUMEN

En la presente tesis se ha investigado el potencial de la cromatografía líquida acoplada a la espectrometría de masas de alta resolución en aproximaciones analíticas tanto de tipo target (análisis de compuestos diana) como non-target (análisis no dirigido). Las ventajas que este tipo de instrumentación ofrece, derivadas de la adquisición del espectro completo con alta resolución y medidas de masa exacta, la hacen idónea para el screening de contaminantes orgánicos así como para estudios metabolómicos tanto de denominación de origen de alimentos como de carácter biomédico.

El trabajo presentado se ha estructurado en tres grandes bloques. En el primero de ellos se aborda el desarrollo y optimización de un método de screening "universal" de contaminantes orgánicos usando un acoplamiento instrumental LC-QTOF MS. Con este fin, se ha desarrollado una base de datos de compuestos, se ha optimizado la metodología analítica y, finalmente, se ha validado dicha metodología en matrices medioambientales. El segundo bloque muestra el potencial de la técnica para la investigación y diagnosis en exhalados pulmonares, los cuales presentan un bajísima concentración de metabolitos. En el tercer y último bloque, se lleva a cabo la búsqueda de compuestos que puedan utilizarse como marcadores de D.O. en muestras de naranja y vino, centrándose especialmente en aquellas pertenecientes a la Comunidad Valenciana.

El primer bloque se inicia con la optimización de los parámetros instrumentales que puedan tener un efecto notorio en la creación de la base de datos y/o en el análisis de las muestras. Para ello, se han investigado los factores que afectan a la sensibilidad instrumental y a la exactitud de masa. Además, se estudia la fragmentación de los compuestos incluidos en la base de datos con el fin de facilitar de forma automática y simultánea la confirmación de la identidad de los compuestos detectados en las muestras. Finalmente se ha evaluado la influencia de la resolución de masa y cromatográfica en la exactitud de masa en el caso de matrices complejas. Posteriormente, se han creado dos bases de datos de contaminantes orgánicos: una

teórica, la cual contiene alrededor de 1000 contaminantes orgánicos reportados en la literatura como analizables por LC-MS. Esta base de datos incluye tiempo de retención y fórmula empírica de fragmentos y aductos de aquellos compuestos de los que se dispone de patrones de referencia y que han sido inyectados de acuerdo con los parámetros previamente optimizados. La segunda base de datos se trata de una librería de espectros experimental que contiene los espectros a alta y baja energía de colisión (estos últimos proporcionando información sobre la fragmentación), así como tiempos de retención, de los compuestos con patrón disponible. Finalmente, se ha evaluado la eficacia de la librería en el análisis de muestras reales.

En un segundo trabajo se evalúa el potencial de los dos principales procedimientos para la investigación de contaminantes orgánicos mediante métodos de screening: non-target y post-target. Para ello, se lleva a cabo un screening de muestras ambientales, alimentos y muestras de interés toxicológico mediante las dos metodologías y se comparan los resultados obtenidos. La primera aproximación (non-target), basada en la deconvolución del cromatograma para la búsqueda de componentes en la muestra, ha demostrado una notable dependencia de la intensidad del pico cromatográfico debido a la baja eficiencia del algoritmo de deconvolución. En el caso de la aproximación non-target, tras la búsqueda de los componentes, los espectros correspondientes son automáticamente comparados con la librería de espectros experimental creada anteriormente, así como con una librería teórica generada a partir de los compuestos incluidos en la base de datos de contaminantes. En cuanto a la segunda aproximación, tipo post-target, esta se basa en la búsqueda, después de la inyección de las muestras, de compuestos seleccionados (target) incluidos en la base de datos. Como se ha indicado anteriormente, ésta incluye información de fragmentación y tiempo de retención de aquellos compuestos con patrón de referencia disponible, es decir aquellos que están también presentes en la librería experimental de espectros. A la vista de los resultados, se concluye que la aproximación post-target resulta la más ventajosa para abordar un screening "universal" de un elevado número de compuestos. Además, los procesos de revisión de datos y los tiempos de procesamiento se reducen considerablemente. Sin embargo, la metodología non-target presenta una excelente capacidad de

confirmación de la identidad de los contaminantes encontrados ya que facilita la comparación de los espectros de fragmentación de patrones con los obtenidos en la muestra. Mediante la aproximación post-target, se encontró un importante número de contaminantes en muestras ambientales y alimentarias, así como drogas de abuso y fármacos en las muestras de orina de voluntarios en tratamientos de desintoxicación.

El primer bloque de la tesis finaliza con una validación cualitativa de la metodología desarrollada en muestras de agua subterránea, superficial y el efluente de una planta de tratamiento de aguas residuales. Se han evaluado dos tipos de relleno en la extracción en fase sólida aplicada a las muestras: Oasis HLB y MCX. El primero ha resultado más genérico, perdiéndose únicamente el fármaco Gabapentina en dicho proceso de preconcentración. Para la validación cualitativa, se fortifican 3 muestras independientes de cada tipo de agua analizada a dos niveles de concentración (0.1 y 1 µg/L) y se comprueba la capacidad del método para detectar (típicamente, usando la molécula protonada) e identificar (mediante al menos dos iones: molécula protonada y un fragmento) los contaminantes seleccionados como modelo: 146 compuestos entre los que se incluyen 52 pesticidas, 52 medicamentos (21 antibióticos), 13 drogas de abuso, 11 hormonas, 11 micotoxinas y 7 agentes de protección UV. El método desarrollado permite la detección de la gran mayoría de los compuestos ensayados y la identificación de un buen número de ellos. Posteriormente, se ha aplicado dicha metodología al análisis de muestras reales, identificando varios de los contaminantes seleccionados, incluso a niveles de concentración inferiores al más bajo validado. También ha sido posible la detección e identificación tentativa de varios contaminantes no incluidos en la validación del método, incluso sin patrón de referencia disponible, gracias a la valiosa información suministrada por el analizador QTOF-MS.

En el segundo bloque se aprovechan las ventajas que ofrece el acoplamiento LC-MS, y en especial HRMS, para aplicaciones metabolómicas que requieren de una elevada sensibilidad instrumental. Este es el caso de los condensados de exhalados pulmonares, cuyas concentraciones de metabolitos son extremadamente bajas y, por tanto, requieren de técnicas más sensibles como son las LC-HRMS frente al típico análisis mediante NMR. En el trabajo realizado, en colaboración con el Instituto de

Estudios Biofuncionales de la Universidad Complutense de Madrid (UCM, Madrid, Spain) y llevado a cabo en el Department of Biomolecular Medicine en Imperial College London (London, UK), se realiza un estudio preliminar sobre las capacidades de LC-HRMS y NMR para abordar la diferenciación de pacientes sanos de aquellos con enfermedades respiratorias, concretamente con obstrucción pulmonar crónica, mediante el análisis no invasivo de condensados de exhalados pulmonares. En base a los resultados obtenidos, se propone el uso de LC-HRMS como aproximación metabolómica estándar para este tipo de análisis.

En el tercer y último bloque se investigan los marcadores que permiten la diferenciación de alimentos según su origen, especialmente en productos de interés para la Comunidad Valenciana y cuya calidad está directamente relacionada con su D.O. El primero de los trabajos incluidos en este capítulo se centra en el análisis de las diferencias a nivel químico presentes en naranjas de diferentes orígenes. Para ello, se han seleccionado muestras de la Comunidad Valenciana y de países del hemisferio sur (Sudáfrica y Argentina) de variedades de maduración tardía. Las muestras completas (piel y pulpa) se trituran, homogenizan y extraen con una mezcla agua:metanol, se diluyen con agua y finalmente se analizan mediante UHPLC-HRMS. Posteriormente, se procesan los datos mediante análisis multivariante para encontrar los que permitirían la distinción en función del origen. Finalmente, se ha ampliado el muestreo a la siguiente campaña, incluyendo también muestras de Brasil, con la finalidad de comprobar la validez de los marcadores encontrados y su posible aplicación a otros destinos. Se ha encontrado un marcador idóneo en ambos casos (primer y segundo muestreo) que corresponde a un compuesto de tiempo de retención 4.83 minutos, que finalmente y gracias a la información espectral en masa exacta ofrecida por LC-HRMS se ha identificado como Citrusin D.

Finalmente, en el último trabajo presentado se ha realizado un estudio similar al de las naranjas. El vino es un producto muy preciado cuyo valor está extremadamente ligado a su origen, entre otros. En este artículo científico se realiza un comparación interlaboratorio para el estudio metabolómico de muestras de vino procedentes de tres importantes denominaciones de origen en España (Ribera del Duero, Rioja y Penedés). Los resultados de ambas plataformas ofrecen una clasificación óptima de

las muestras en base a marcadores definidos en las rutas metabólicas de la vitis vinífera. Además, se discute sobre las ventajas y desventajas de ambas plataformas, no solo en el aspecto instrumental (TOF vs Orbitrap) sino también en el procesamiento de los datos y la selección de los marcadores. Como cabía esperar por ser geográficamente la más diversa, la denomiancion de origen Penedés era separada de las otras dos mediante un simple análisis no-supervisado PCA. Para separar completamente las 3 D.O. se analizaron mediante PLS-DA obteniéndose una clasificación correcta para todas las muestras. Se encontraron diversos marcadores de la familia de los polifenoles como la Catequina, Epicatequina y Galactocatequina, entre otros. Finalmente, los marcadores de cada plataforma fueron cuantificados en un modo target en la otra plataforma. Se demostró que, para ambas plataformas, los diferentes marcadores eran significativos y que por tanto, el tratamiento de datos había filtrado estos marcadores. El modelo estadístico aplicado haciendo uso exclusivamente de los marcadores fue capaz de separar perfectamente las diferentes denominaciones de origen mediante PCA.

SUMMARY

In this thesis, the potential of the liquid chromatography coupled to high resolution mass spectrometry has been investigated in non-target and post-target approaches. This type of instrumentation offers some advantages like full-spectrum acquisition and high-resolution/accurate mass measurements that make it ideal for wide-scope screening of organic pollutants as well as for metabolomic studies either in food origin determination or in biomedical research.

This work is divided in three main parts. In the first one, the optimization of a "universal" (liquid chromatography suitable) wide-scope screening of organic pollutants is performed by LC-QTOF MS. To this purpose, a contaminant database has been created, the analytical methodology optimized and, finally, validated in environmental water samples. The second part shows the potential of this technique for the investigation and diagnosis in human breathe exhaled condensates, which are characterized by extremely low metabolite concentrations. In the third and last part, a biomarker search has been carried out in order to use them as a tool for food origin determination, concretely in orange and wine samples which are of especial interest in the Valencia region.

The first section begins with the optimization of the instrumental parameters with a notorious effect in the creation of a compound database and/or in the sample analysis. Thus, factors that could affect the sensitivity and mass accuracy have been investigated. Furthermore, fragmentation of the compounds included in the database has been studied in order to perform an automatic and simultaneous confirmation of the identity of these compounds. Finally, the influence of the mass and chromatographic resolution in the mass accuracy was evaluated for complex matrices like vegetables or wastewater. Afterwards, two organic pollutants databases were built: a theoretical one with approximately 1000 LC-Ms suitable contaminants reported in the literature that included retention time and empirical formula of the fragments/adducts of those compounds with available reference standard (which were previously injected

within the optimized parameters). The second database is an empirical library which contain mass spectra (as well as retention time information and promoted fragmentation spectra) of the available reference standard compounds. Finally, efficiency of the empirical library in real samples application was tested.

In the next work, the potential of the main approaches for the contaminant searching –non-target and post-target- was evaluated. For this purpose, environmental, food and toxicological sample screening was carried out by these two methodologies and results obtained were compared. The first approach (the non-target one), based on the chromatogram deconvolution to find the components present in the sample, demonstrated high dependence on the intensity of the chromatographic peak due to the low efficiency of the deconvolution algorithm. In this case, the spectra of the peaks found were automatically compared with those included in a home-made empirical mass spectra library previously built as well as a theoretical one from the compounds included in the pollutants data-base. Regarding the post-target approach, it was based on the post-injection search of the contaminants present in the previously mentioned contaminants database that, as stated before, included fragmentation and retention time information for those compounds with available reference standard, i.e. those present in the empirical mass spectra library. Results revealed that post-target approach was most advantageous for the "universal" organic pollutants screening due to the lower number of false negatives. Furthermore, it permitted to considerably reduce the processing and results reviewing time. However, the non-target methodology showed excellent skills for the confirmation of the identity of the contaminants found due to it facilitates the HE mass spectra comparison between sample and standard. Using the post-target method, a significant number of contaminants in the environmental and food samples and several pharmaceuticals and drugs of abuse positives in the urine samples was found.

The first section ends with the performance of a qualitative validation of the previously developed methodology in environmental water samples, groundwater, surface water and effluent wastewater. Two pre-concentration solid phase extraction methods with two different cartridges -Oasis HLB and MCX- were tested. The HLB one resulted more generic, with the only loss of the pharmaceutical gabapentin during the

extraction process. Then, 3 different matrices from each type (getting a total of 9 samples) were spiked at two concentration levels (0.1 y 1 µg/L). The ability of the methodology to detect and identify the contaminants belonging to different chemical families was evaluated: 146 compounds including 52 pesticides, 52 pharmaceuticals (21 antibiotics), 13 drugs of abuse, 11 hormones, 11 mycotoxins and 7 UV protection agents. The screening method allowed the detection of most of the compounds, at least, at the highest concentration as well as the reliable identification of the majority of them. Finally, this screening method was applied to real environmental samples, identifying some of the compounds included in the validation, even at concentrations well-below to the lowest concentration level tested. It was also possible to detect and identify compounds with no reference standard available thanks to the capabilities of the TOF analyzers.

The second part of the thesis is focused on the study of the advantages offered by the LC-MS system, especially with HRMS instrumentation, to perform specific metabolomic approaches that require a high instrumental sensitivity. This is the case of the exhaled breath condensates (EBC) which metabolite concentrations are extremely low and require very sensitive techniques, such as LC-HRMS. In this short communication, a preliminary study about capabilities of LC-HRMS and typically used NMR to face to the differentiation of healthy patients and patients with airway disease (chronic obstructive pulmonary disease) by the analysis of non-invasive EBC has been carried out in collaboration with the Instituto de Estudios Biofuncionales de la Universidad Complutense de Madrid (UCM, Madrid, Spain) and performed in the Department of Biomolecular Medicine at Imperial College London (London, UK). In conclusion, UHPLC-HRMS is proposed as a standard metabolomic approach in the analysis of EBC.

In the third and last section of this thesis, biomarkers for the unequivocal distinction of certain properties of food and beverages are evaluated, such as geographical origin. This study is focused especially on those with a socio-economic interest in the Comunidad Valenciana. The first work in this chapter is based on the chemical differences of different origins of oranges. To this aim, late-maturing orange samples from Valencia region and from south hemisphere countries (South-Africa and

Argentina) were analysed by UHPLC-HRMS. Data were processed using multivariate analysis to find the most adequate markers to distinguish the oranges origin. Then, the sample number was increased with two different seasons oranges and included additional origin (Brazil) to check the validity of the markers found. Tentative identification was performed to the most significant marker and was tentatively identified as Citrusin D.

Finally, in the last work, a similar study was performed for wine samples. Wine is a relatively high cost product which value is extremely linked to its geographical origin, amongst others. In this paper, an interlaboratory comparison has been performed for the metabolomic study of wine samples from three important Spanish wine production areas (Ribera del Duero, Rioja and Penedés). Results from both platforms showed an optimal origin classification based on markers well defined in vitis vinifera metabolic pathways. Moreover, advantages and drawbacks of both platforms are extensively discussed, not just in terms of instrument used but also based on the different data treatment and markers selection. As expected due to their higher geographical difference, samples from Penedes were easily separated by analysing the data using unsupervised PCA. In order to completely separate the 3 D.O., data were analysed by PLS-DA, obtaining a correct classification for all the samples. Several polyphenol compounds were found as biomarkers such as Catechi, Epicatechin and Gallactocatechin, amongst others. Finally, markes found in each platform were screened in a target way in the other one. It was demonstrated that, for both platforms, the selected metabolites were significant and then the data treatment was filtering them. The statistical model applied using exclusively the biomarkers selected allowed the perfect separation of the 3 D.O. studied by PCA.

ÍNDICE GENERAL

pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry"

Rapid Commun. Mass Spectrom., 25 (2011) 355–369

"Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF MS"

Anal. Methods, 4 (2012) 196–209

Qualitative validation of LC-(Q)TOF MS screening method for organic pollutants in waters"

J. Chromatogr. A, 1276 (2013) 47–57

CAPÍTULO 3. METABOLÓMICA EN EXHALADOS PULMONARES HUMANOS

"Is NMR-based metabonomic analysis of Exhaled Breath Condensate accurate?"

Eur Respir J., 37 (2011) 468-470

CAPÍTULO 4: APROXIMACIONES METABOLÓMICAS PARA LA AUTENTIFICACIÓN DE LA DENOMINACIÓN DE ORIGEN DE ALIMENTOS

GENERAL INDEX

"Building an empirical mass spectra library for screening of organic

pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry"

Rapid Commun. Mass Spectrom., 25 (2011) 355–369

"Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF MS"

Anal. Methods, 4 (2012) 196–209

Qualitative validation of LC-(Q)TOF MS screening method for organic pollutants in waters"

J. Chromatogr. A, 1276 (2013) 47–57

CAPÍTULO 3. HRMS METABOLOMICS IN HUMAN BREATH EXHALATED

"Is NMR-based metabonomic analysis of Exhaled Breath Condensate accurate?"

CHAPTER 4: METABOLOMIC APPROACHES FOR FOOD ORIGIN AUTHENTICATION

(FOODOMICS)

Índice de acrónimos/Abbreviations

ACN	Acetonitrile
APCI	Atmospheric Pressure Chemical Ionisation
API	Atmospheric Pressure Interface
APPI	Atmospheric Pressure Photoionisation
CID	Collision Induced Dissociation
DART	Direct Analysis in Real Time
D.O.	Denominación de Origen/Denomination of Origin
EDC	Enhanced Duty Cycle
EM	Efecto matriz/Matriz Effect
EPA	United States Environmental Protection Agency
ESI	Electrospray Ionisation
ESI-MS	Mass Spectrometry with Electrospray Ionisation
EWW	Effluent wastewater
FT-ICR	Fourier Transform Ion Cyclotron Resonance
FWHM	Full-Width Half Maximum
GC-EI-MS	Gas Chromatography coupled to Electron Ionisation Mass Spectrometry
GC-MS	Gas Chromatography coupled to Mass Spectrometry
HCOOH	Formic acid
HILIC	Hydrophilic Interaction Liquid Chromatography
HPLC	High-Performance Liquid Chroamtography
HR	High Resolution
HRMS	High Resolution Mass Spectrometry
IPs	Identification Points
IT	Ion Trap analyser
IWW	Influent Wastewater
LC	Liquid Chroamtography
LC-ESI-MS/MS	Liquid Chromatography coupled to Electrospray Ionisation Mass Spectrometry
LC-MS	Liquid Chromatography coupled to Mass Spectrometry

LC-MS/MS	Liquid Chromatography coupled to tandem Mass Spectrometry
LLE	Liquid-Liquid Extraction
LOD	Limit of Detection
LOI	Limit of Identification
LR	Low Resolution
LRMS	Low Resolution Mass Spectrometry
MALDI	Matrix Assisted Laser Desorption/Ionisation
MeOH	Methanol
MS/MS	Tandem Mass Spectrometry
m/z	Mass to charge ratio
NMR	Nuclear Magnetic Resonance
NPLC	Normal-Phase Liquid Chromatography
nw-XIC	narrow mass window eXtracted Ion Chromatogram
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
PC	Principal Component
PCA	Principal Components Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
ppm	Parts per million
ppb	Parts por billion
Q	Quadrupole analyser
QC	Quality Control
QLIT	Hybrid Quadrupole-Linear Ion Trap analyser
QTOF	Hybrid Quadrupole-Time Of Flight analyser
QqQ	Triple Quadrupole analyser
RPLC	Reverse-Phase Liquid Chromatography
RSD	Relative Standard Deviation
SDL	Screening Detection Limit
SIM	Single Ion Monitoring
S/N	Signal to Noise Ratio
SPE	Solid Phase Extraction

SRM	Single Reaction Monitoring
TDC	Time-to-Digital converter
TIC	Total Ion Current
TPs	Transformation Products
UHPLC	Ultra High-Performance Liquid Chroamtography
UV	Ultraviolet Detection
(nw-)XIC	(narrow window) eXtracted Ion Chromatogram

Objetivos y plan de trabajo

OBJETIVOS

En esta tesis se persiguen **dos objetivos generales**, sobre la base de las capacidades analíticas del potente acoplamiento instrumental cromatografía líquida-espectrometría de masas de alta resolución (LC-HRMS), en concreto con analizador híbrido QTOF MS. Por un lado, se pretende desarrollar una metodología analítica avanzada de tipo "universal" para el screening de un elevado número de contaminantes orgánicos pertenecientes a diferentes familias (pesticidas, fármacos, drogas de abuso, productos de cuidado personal, etc.) en muestras de interés ambiental, alimenticio y toxicológico. Por otro lado, se pretende investigar el potencial de este acoplamiento instrumental para llevar a cabo estudios metabolómicos en muestras de alimentos, con el objetivo de encontrar marcadores distintivos de una o más características (como por ejemplo, el origen geográfico de los alimentos). También se pretende explorar la capacidad de LC-QTOF MS (en comparación con la típicamente utilizada NMR) en aplicaciones metabolómicas que requieren de una elevada sensibilidad, como es el caso de los condensados de exhalados pulmonares (los cuales presentan la ventaja de ser muestras no invasivas), concretamente para la clasificación de pacientes con obstrucción pulmonar crónica.

Para alcanzar el objetivo general relacionado con el **screening de contaminantes orgánicos**, se han establecido una serie de **objetivos específicos**:

1. Creación de una librería experimental de espectros de masas a partir de patrones disponibles de contaminantes ampliamente detectados en muestras reales bajo condiciones óptimas de análisis, que deberán ser establecidas previamente.

2. Creación de una amplia base de datos teórica de compuestos analizables por LC-MS a partir de información bibliográfica.

3. Elucidación estructural de los fragmentos de compuestos incluidos en la librería de espectros experimental para facilitar la identificación de los contaminantes en el análisis posterior de las muestras.

4. Comparación de las aproximaciones post-target y non-target para el screening "universal" de contaminantes orgánicos en diferentes matrices de interés.

5. Aplicación de las metodologías desarrolladas al análisis de muestras reales.

6. Evaluación de su potencial para la detección e identificación de los contaminantes encontrados en las muestras, así como para la identificación tentativa de otros contaminantes cuyo patrón no esté disponible en el laboratorio.

7. Validación cualitativa de la metodología elegida para compuestos seleccionados, usados como modelo.

Por otro lado, en relación con el **estudio metabolómico**, se establecen los siguientes **objetivos**:

1. Explorar el potencial del acoplamiento LC-QTOF MS para el análisis no-dirigido de muestras, con especial atención a los distintos mecanismos cromatográficos de separación y a las características de la espectrometría de masas frente a otras técnicas típicamente utilizadas.

2. Comparar distintos algoritmos de pre-procesamiento de datos para la obtención de la máxima información posible (componentes dentro del cromatograma) a partir de los datos obtenidos en el análisis de las muestras.

3. Obtener modelos estadísticos que permitan la interpretación y diferenciación de las muestras según su origen, variedad, etc., así como la obtención de los principales marcadores que permitan distinguirlas.

4. Elucidar la estructura, identificar e interpretar la significancia de los biomarcadores encontrados.

5. Validar los resultados obtenidos mediante la ampliación de las muestras analizadas.

Más específicamente, para los estudios metabolómicos realizados se pretende:

6. Evaluar la aplicabilidad y eficiencia de las técnicas espectrométricas para llevar a cabo estudios metabolómicos en muestras de elevada dificultad por su baja concentración de metabolitos como son los condesados de exhalados pulmonares, comparando dicha aproximación con la NMR, la cual está mucho más consolidada pero presenta desventajas en cuanto a su sensibilidad.

7. Encontrar un/os marcador/es fiable/s para la determinación del origen de naranjas, con especial atención a frutos provenientes de la región de Valencia, pudiendo así distinguirlos inequívocamente de aquellos procedentes del hemisferio sur (Sudamérica, Sudáfrica).

8. Estudiar las diferencias en la concentración de los metabolitos de vinos españoles en base a diferentes variables como el tipo de uva, la región donde se produce o el año de producción. Asimismo, intentar hallar marcadores que permitan diferenciar los diferentes vinos.

PLAN DE TRABAJO

Para alcanzar los objetivos indicados, se establecieron dos planes de trabajo dependiendo de la aproximación a considerar. Así pues, para el screening de contaminantes orgánicos el plan de trabajo fue:

1. Optimización de las condiciones cromatográficas y de la espectrometría de masas. Para ello, se optimizarán la resolución cromatográfica, así como, la exactitud de masa y la sensibilidad del TOF MS.

2. Estudio de las condiciones óptimas para la fragmentación de los compuestos con el fin de facilitar su identificación.

3. Inyección del máximo número de patrones de referencia posibles bajo las condiciones experimentales previamente optimizadas con el fin de incluir dicha información en la librería de espectros.

4. Búsqueda bibliográfica de contaminantes orgánicos analizables mediante LC-MS bajo condiciones cromatográficas y espectrométricas similares a las optimizadas.

5. Selección de la aproximación (post-target y/o non-target) más idónea para el screening de muestras reales y desarrollo de dicha aproximación en diferentes matrices de interés alimentario, toxicológico y medioambiental.

6. Selección de los compuestos y los niveles de concentración representativos para llevar a cabo la validación cualitativa con el fin de evaluar el potencial de la metodología desarrollada.

Por otro lado, en el caso de los estudios metabolómicos el plan de trabajo fue el siguiente:

1. Estudio de los diferentes tratamientos de muestra con el fin de cubrir la mayor cantidad de metabolitos posibles.

2. Estudio de las condiciones cromatográficas óptimas para los diferentes tratamientos propuestos.

3. Análisis de las muestras de forma aleatoria y evitando posibles errores sistemáticos.

4. Estudio de las diferentes variables y aproximaciones para el pre-procesamiento de datos, es decir la búsqueda de los metabolitos mediante aproximación non-target.

5. Estudio de la varianza de los datos mediante técnicas estadísticas multivariante, y más concretamente mediante PCA.

6. Desarrollo de modelos estadísticos dirigidos; tanto univariantes (ANOVA) como multivariantes (PLS, PLS-DA, OPLS-DA), para la determinación de las diferencias entre muestras.

7. Determinación y elucidación de los marcadores óptimos para los diferentes propósitos

8. Confirmación, en los casos posibles, de la identidad de los marcadores elucidados así como de la validez del modelo generado y de dichos marcadores.

Objectives and working plan

OBJECTIVES

This thesis pursues the accomplishment of **two main different objectives** based on the study of the analytical capabilities of coupling liquid chromatography and high resolution mass spectrometry (LC-HRMS). Firstly, it will be performed an "universal" wide-scope screening of organic pollutants belonging to different families (pesticides, pharmaceuticals, drugs of abuse, personal care products, etc.) in different matrices with environmental, food and toxicological interest.

On the other hand, the potential of LC-HRMS will be investigated to carry out metabolomic studies in food samples with the objective to find distinctive biomarkers for certain qualities like geographical origin. Previously, the potential of LC-QTOF MS (compared to typically used NMR techinique) in applications that require high sensitivity will be explored, such as the exhaled breath condensates, which show the advantage of being a non-invase sample, but have a low concentration of metabolites. Concretely, it will be applied to the diferentition of subjects with pulmonary disease like chronic obstructive pulmonary disease (COPD).

To reach the general goal regarding the wide-scope screening of organic pollutants development, more **specific objectives** are established:

1. Optimization of the chromatographic and mass spectrometry conditions. Thus, chromatographic resolution, as well as mass accuracy and sensitivity of TOF MS will be optimized.

2. Study of the optimal conditions for the fragmentation of the contaminants with the aim of facilitating its identification.

3. Development of an empirical mass spectra library including the available reference standard compounds under previously optimized instrumental conditions.

4. Building a theoretical pollutants database with LC-MS amenable pollutants according to literature information.

5. Structural elucidation of the fragment ions obtained from those compounds included in the empirical mass spectra library to facilitate the identification of the contaminants.

6. Comparison of post-target and non-target approaches for the wide-scope screening of organic contaminants in the different matrices studied.

7. Study of the applicability of the methodology/es applied for the real sample analysis.

8. Evaluation of the potential of the instrumental coupling used for the detection and identification of the pollutants found, as well as, for those present in the samples analyzed which reference standard is unavailable in our laboratory.

9. Qualitatively validate the selected methodology for the available analytes.

Regarding the metabolomic study, following **objectives** are established:

1. Explore the ability of the instrumental coupling above mentioned for the completely untargeted analysis of the samples, with especial attention to the different chromatographic separation mechanisms and to the characteristics of mass spectrometry against other typically used techniques.

2. Compare different data pre-processing algorithms to achieve the maximum information from the chromatograms.

3. Obtain statistical models that allow the interpretation and differentiation of the samples as a function of the geographical origin, fruit variety, etc., as well as, that permit getting the main biomarkers for this differentiation.

4. Structural elucidation, identification and interpretation of the biomarkers found.

5. Validate the results by increasing the number of samples analyzed.

More specifically, the goal of these metabolomic studies is:

6. Evaluate the efficiency and applicability of the spectrometric techniques to carry out metabolomic studies on highly difficult matrices due to their low metabolite concentration such as exhaled lung condensates. Compare the results with those obtained by NMR, typically the most used technique.

7. Find one or more reliable biomarkers for the determination of the oranges origin with especial attention to those from the Valencia region, in order to unequivocally differentiate them from samples from south hemisphere (South-America and South-Africa).

8. Study the differences in Spanish wines metabolite concentration based on different variables such as grape variety, geographical origin or production year. Moreover, find biomarkers that permit to distinguish the wines.

WORKING PLAN

To reach the goals mentioned above, two different working plans were established depending on the approach considered. Thus, in the case of the organic pollutants screening the plan to follow was:

1. Optimization of the chromatographic and spectrometric conditions. Factors to have into account are those influencing the chromatographic resolution as well as TOF MS mass accuracy and sensitivity.

2. Study of the optimal fragmentation conditions with the aim to facilitate the compound identification.

3. Injection, under the optimal conditions, of the highest possible number of reference standards to include the information in the mass spectral library.

4. Literature search of LC amenable organic contaminants to be included in the theoretical database.

5. Selection of the most appropriate approach (post-target and/or non-target) for the wide-scope screening of real environmental, food and biological samples.

6. Selection of the representative compounds and concentration levels in order to perform a reliable qualitative validation with the final aim to evaluate the potential of the developed methodology.

On the other hand, in the case of the metabolomic studies the working plan was:

1. Study of the different simple treatments to cover the highest number of compounds possible.

2. Investigate the optimal chromatographic conditions for the different sample treatments proposed.

3. Analysis of the samples in a random way and avoiding systematic errors.

4. Study of the different variables and approaches for the data pre-processing, i.e. metabolites peak picking in an untargeted way.

5. Study of the data variance by multivariate statistical tools, and more concretely by PCA.

6. Development of the statistical biased models: both univariate (ANOVA) and multivariate (PLS, PLS-DA, OPLS-DA) for the classification of the samples.

7. Determination and elucidation of the main biomarkers for the different targets.

8. Confirmation, when possible, of the identity of the elucidated markers and the applicability of the model.

CAPÍTULO 1
Introducción general

CAPÍTULO 1. INTRODUCCIÓN GENERAL

La sociedad actual, en especial los países más desarrollados, está sufriendo un continuo cambio en cuanto al estilo de vida. Esto implica, en general, una mejora del nivel de vida así como de la esperanza de vida. Sin embargo, este acelerado cambio puede tener sus efectos secundarios (sobre todo a largo plazo) en los seres humanos y demás especies animales y vegetales.

El número de compuestos químicos que se aplican diariamente crece de forma drástica (Portolés, Pitarch et al. 2011). Debido a su cada vez más amplio uso, estos compuestos llegan a representar un riesgo medioambiental y de seguridad alimentaria, por lo que existe preocupación creciente por parte de distintos organismos internacionales. Hablamos de compuestos que se utilizan ampliamente en la vida cotidiana, como fármacos, productos de cuidado personal o productos de limpieza, pero también otros compuestos como conservantes alimenticios, retardantes de llama, pesticidas, etc; los cuales son cada vez más necesitados debido a la creciente demanda (Richardson and Ternes 2014).

Esto puede conllevar problemas medioambientales y de seguridad alimentaria y, por tanto, se requiere una actuación acorde con dicho problema. Aunque recientemente la disponibilidad de técnicas instrumentales avanzadas ha posibilitado el desarrollo de métodos analíticos muy fiables y rápidos para la detección de un número considerable de contaminantes, debemos dar un paso adelante en cuanto al análisis de contaminantes se requiere. Debemos anticiparnos al problema y al uso de compuestos similares.

Por otro lado, los nuevos tiempos han propiciado una globalización, que afecta tanto a la información y la economía como a diferentes productos. No es extraño ya ver vegetales y frutas típicos de zonas tropicales, incluso en los meses invernales. Esto ha propiciado la exaltación de las llamadas denominaciones de origen, las cuales proclaman la calidad de sus productos en base a su origen y/o tratamiento. Cada

denominación de origen define distintos parámetros de calidad, no sólo de delimitación geográfica sino también en cuanto a variedades utilizadas, tiempos de cosecha, tratamientos permitidos y, en algunos casos como en los vinos, en tratamiento y manipulación para conseguir el producto final. El objetivo de las denominaciones de origen es ofrecer al consumidor una garantía de calidad frente a los diferentes productos disponibles a disposición de los usuarios. Sin embargo, nuevos métodos analíticos son necesarios para asegurar dichas características de calidad.

En esta tesis doctoral se pretenden investigar estas problemáticas. Por un lado, se procura avanzar en el análisis de contaminantes mediante el desarrollo de métodos de screening basados en el uso del acoplamiento LC-HRMS, el cual facilita la detección e identificación fiable de contaminantes orgánicos en muestras medioambientales y alimentarias. Por otro, se intenta avanzar en la problemática de las denominaciones de origen ofreciendo soluciones a posibles fraudes.

En este capítulo de Introduccion, se comenta en primer lugar el acoplamiento LC-HRMS y sus principales características, que lo convierten en una herramienta poderosa para fines de screening. Posteriormente, se discuten las aproximaciones target y non-target en análisis de contaminantes orgánicos y residuos en general, para finalizar con una breve introducción al tratamiento de muestra, desde el punto de visto de los análisis llevados a cabo en esta Tesis Doctoral.

1.1. CROMATOGRAFÍA LÍQUIDA ACOPLADA A ESPECTROMETRÍA DE MASAS DE ALTA RESOLUCIÓN (LC-HRMS)

Para enfrentarse a la problemática de los dos principales aspectos tratados en esta Tesis, screening de nuevos contaminantes y metobolómica, se requieren técnicas analíticas avanzadas, capaces de ofrecer sensibilidad, elevada información estructural, "universalidad" y/o versatilidad. Obviamente, no existe una herramienta única capaz de generar todas las expectativas requeridas, pues las diferentes propuestas que puedan ajustarse al perfil exigido presentan siempre ciertos inconvenientes. Sin embargo, la técnica por excelencia en el campo del análisis de contaminantes orgánicos es, sin duda, la cromatografía acoplada a espectrometría de masas. Entre las diferentes posibilidades que se ofrecen, resulta esencial la espectrometría de masas de alta resolución para afrontar dichos objetivos de tipo screening.

Otras técnicas pueden presentar una sensibilidad similar, lo cual es un requisito debido a las bajas concentraciones en las que suelen estar los contaminantes en las diferentes matrices, como por ejemplo LC-Fluorescencia, GC-captura de electrones, etc. Sin embargo, ninguna de estas técnicas ofrece la selectividad necesaria ni aporta información estructural para afrontar un screening de amplio espectro de compuestos orgánicos.

En lo que al campo de la metabolómica se refiere, la elección del acoplamiento LC-MS o GC-MS no es tan ventajosa como en el campo de residuos y/o medioambiental, pues la sensibilidad, aunque importante, no juega un papel tan crucial. Por ello, las técnicas de resonancia magnética nuclear han sido las más utilizadas por su robustez y elevada información estructural, lo que facilita la creación de bases de datos y, por tanto la elucidación de (bio)marcadores. No obstante, existe un creciente interés en la espectrometría de masas en cuanto a la metabolómica se refiere, lo que se deduce a partir del creciente número de publicaciones sobre investigaciones en este campo (Wishart 2008).

1.1.1. Cromatografía Líquida

A pesar de la elevada selectividad que proporciona la detección por espectrometría de masas, muy especialmente los analizadores de masas de elevada resolución, resulta necesario realizar una separación previa de los diferentes componentes de la muestra con el fin de aislar en lo posible los analitos para facilitar su identificación y cuantificación, evitando o minimizando de esta manera interferencias isobáricas o efectos matriz. La técnica de separación más utilizada en combinación con MS es la cromatografía, que se basa en la diferente distribución de los componentes de la muestra entre una fase estacionaria y una fase móvil (normalmente un gas o una mezcla de disolventes y sales con distinta fuerza eluotrópica) debido a diferencias de velocidad de desplazamiento de los compuestos al ser arrastrados por la fase móvil a través de la fase estacionaria. En función de la naturaleza de estas dos fases se pueden distinguir diferentes modalidades cromatográficas.

En esta tesis doctoral se ha trabajado exclusivamente con cromatografía líquida (LC), la cual es sin duda la técnica más adecuada para separar compuestos de polaridad media-alta, poco volátiles y térmicamente lábiles. La fase móvil es líquida y fluye a través de una columna empaquetada con una fase estacionaria que contiene un grupo activo (Snyder 2009).

Complementaria a la LC se presenta la cromatografía de gases (GC) cuya fase móvil es un gas inerte (generalmente He). En el caso de la GC, los analitos susceptibles de análisis son poco polares y termo-estables. Existen ventajas y desventajas en el uso de ambas técnicas (LC y GC) pero la elección de una u otra se basa, aparte de la disponibilidad, en las características fício-químicas del analito (Hernández, Ibáñez et al. 2015). En esta tesis se ha trabajado buscando una visión global sobre la composición de la muestra, que no ha estado centrada exclusivamente en el tipo de analitos. Se eligió la LC como técnica de separación, por su mayor universalidad, por la simplicidad del tratamiento de muestra, por su mayor adecuación a analitos polares (como son la mayoría de contaminantes emergentes y productos de

transformación/metabolitos), y por las ventajas que ofrecen las fuentes de ionización usadas en LC-MS, que se verán mas adelante.
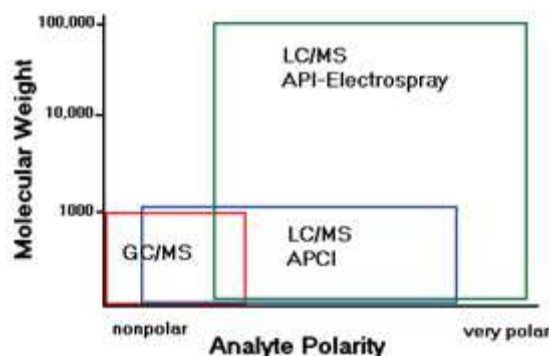
Figura 1.A. Rango de aplicación de los acoplamientos GC-MS y LC-MS más ampliamente utilizados en función de la polaridad y masa molecular de los compuestos.

Como se puede observar en la Figura 1.A., el rango de masa molecular en los compuestos analizables por LC es mucho mayor. Esto, unido a la mayor afinidad para compuestos de polaridad media-alta, hace que LC sea más adecuada para los estudios de tipo screening, sobre todo cuando se pretende incluir los llamados "compuestos emergentes", los cuales son típicamente polares. En los estudios metabolómicos, este hecho es todavía más evidente pues el rango de masas moleculares y la polaridad de los metabolitos ha de ser lo más amplio posible.

A continuación, se comentan brevemente los diferentes mecanismos de separación en LC utilizados durante esta tesis:

Cromatografía liquida en Fase Reversa (RPLC)

La LC en fase reversa es sin lugar a dudas la técnica más utilizada entre las diferentes posibilidades que ofrece la LC debido a su robustez, bajo coste de las columnas y universalidad, ya que permite separar un gran número de compuestos orgánicos de polaridad media-alta. La popularidad de este tipo de cromatografía reside además en la idoneidad para el acoplamiento LC-MS por la volatilidad de los disolventes orgánicos utilizados.

La LC en fase reversa se basa en la retención de los compuestos en una fase estacionaria apolar, los cuales son eluidos gradualmente al aumentar la fuerza eluotrópica de la fase móvil, es decir, al disminuir la polaridad de los disolventes. Así pues, contrariamente a la LC en fase normal, el gradiente de elución se realiza disminuyendo gradualmente la polaridad de la fase móvil. Por tanto, los compuestos más polares eluyen inicialmente, ya que su interacción con la fase estacionaria apolar es prácticamente nula, mientras que los compuestos más apolares se verán más retenidos en la fase estacionaria, requiriendo solventes más apolares (MeOH, ACN, etc.) para vencer dicha interacción.

La fase estacionaria suele estar formada por partículas de sílice modificada químicamente en la que los grupos hidroxilos (-OH) son intercambiados por cadenas alquílicas (C8, C18, etc.), aunque también existen columnas en base polimérica o con partículas Sílice-Carbón. El tipo de cadena alquílica (y por tanto la polaridad) es el factor más importante en la elección de la fase estacionaria aunque también puede influir el tipo de base utilizada, pues de ello dependen en gran medida los rangos o límites de pH permitidos. Las columnas cromatográficas en base sílice, aunque son las de uso más común, no permiten trabajar a pH alcalinos pues a tales pH se produciría la hidrólisis de las cadenas alquílicas unidas a los átomos de sílice.

Los modificadores suelen jugar un papel crucial en LC como reguladores del pH o formadores de pares iónicos, pues la inmensa mayoría de los compuestos analizables por esta técnica contienen uno o varios grupos ácidos o básicos, cuyas características físico-químicas pueden verse alteradas por cambios de pH, dificultando así el correcto desarrollo de la cromatografía (Poole 2003). Por otro lado, en el caso del acoplamiento LC-MS, los modificadores son muy útiles para mejorar la eficiencia de la ionización (especialmente en el caso de las interfase ESI), favoreciendo (o evitando, según convenga) la formación de iones positivos o negativos según contenga la molécula grupos básicos o ácidos, respectivamente. Como norma general, la adición a la fase móvil de pequeñas cantidades de modificadores que cedan o acepten protones aumenta la sensibilidad frente al analito en cuestión pues favorece la formación de iones pseudomoleculares protonados ($[M+H]^+$) o desprotonados ($[M-H]^-$),

y evita la formación de iones generalmente indeseables como los aductos sodiados ([M+Na]+), los cuales producen espectros de fragmentación pobres.

Cabe destacar la importancia de utilizar modificadores volátiles en el acoplamiento LC-MS. El uso de sales inorgánicas, tales como tampones fosfato, que son muy comunes en metodologías HPLC con otro tipo de detectores, provoca la precipitación de éstas en la interfase lo que conlleva a una pérdida paulatina de sensibilidad o incluso al bloqueo completo del cono extractor. Sales o tampones de acetato, formiato e ion amonio son las más aconsejables y utilizadas.

Cromatografía Líquida de Interacción hidrofílica (HILIC)

Conocida por su nombre en inglés HILIC (Hydrophilic Interaction Liquid Chromatography), es una de las modalidades de cromatografía líquida de más reciente aparición aunque su popularidad está en continuo aumento, según refleja el notable incremento de publicaciones relacionadas con esta técnica de separación (Figura 1.B.).
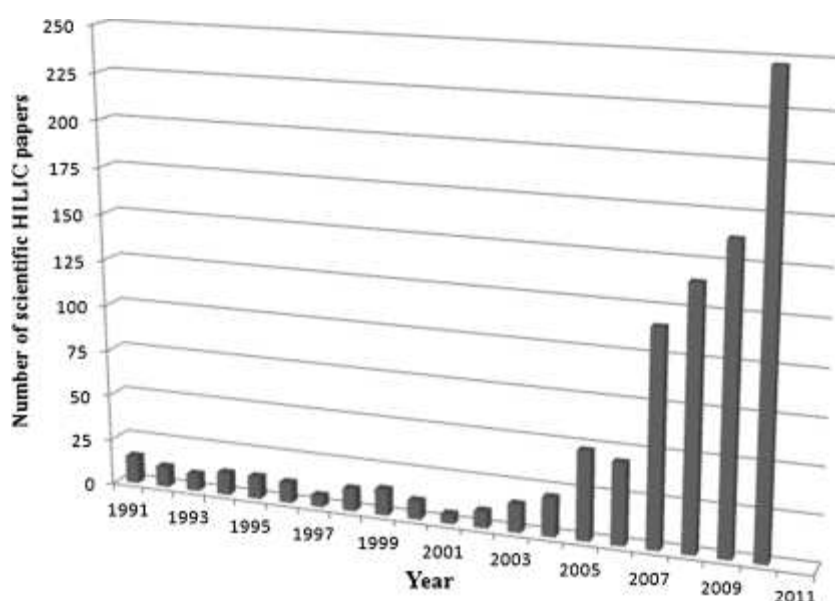
Figura 1.B. Resultados de la busqueda de HILIC en SciFinder Scholar

En realidad, la técnica HILIC no es un nuevo método de separación sino más bien una variación de la LC en fase normal (NPLC). En este tipo de mecanismos, se busca retener los compuestos de elevada polaridad o incluso en su forma iónica, obteniendo así una técnica de separación casi complementaria a la RPLC. Así pues, mientras con columnas RPLC se puede conseguir la separación de analitos de polaridad media-baja, la opción HILIC permite separar compuestos en un rango de polaridad totalmente distinto, favoreciéndose la retención de analitos muy polares y/o iones, como es el caso de distintos los aminoácidos o pequeños péptidos. Estos analitos suelen ser además, moléculas relativamente pequeñas y con marcado carácter acido-base.

Los compuestos separados mediante cromatografía HILIC eluyen en orden de hidrofobicidad decreciente, aunque este parámetro está íntimamente ligado a la polaridad. De hecho, se cree que el mecanismo de acción de las columnas HILIC es principalmente de tipo partición líquido-líquido, creada por la formación de una capa de agua en la superficie de la fase estacionaria. La retención está también influenciada por interacciones de puentes de hidrogeno y fuerzas electroestáticas.

La mayor diferencia entre HILIC y NPLC reside en las distintas fases móviles utilizadas. Tradicionalmente, NPLC utiliza solventes 100% orgánicos o mezclas miscibles de ellos, mientras que la cromatografía HILIC usa fases móviles orgánicas miscibles en agua. Además, es necesario un pequeño contenido inicial de agua para formar la fina capa de agua en la superficie de la fase estacionaria que permite la interacción con el analito. Por tanto, cuando se utilizan gradientes HILIC, se aumenta la fuerza eluotrópica mediante el incremento de la proporción acuosa del solvente de elución. La principal ventaja de HILIC sobre NPLC radica en el tipo de solventes utilizados. Así pues, mientras que ambas técnicas son utilizables para compuestos que eluyen en la zona del volumen muerto en RPLC, la HILIC salva los problemas de solubilidad normalmente encontrados en NPLC en muestras acuosas.

Por otra parte, la utilización de solventes orgánicos tales como ACN favorece el acoplamiento LC-MS, especialmente cuando se hace uso de interfases ESI, pues facilita la solvatación de la fase móvil, mejorando así la sensibilidad.

HILIC se ha establecido como el modo de separación por excelencia para compuestos hidrofílicos o amfifílicos (cargados o no), que resultan demasiado polares para ser correctamente retenidos en RPLC pero que no presentan suficiente carga como para permitir una retención electrostática efectiva en cromatografía de intercambio iónico, la cual es además difícilmente acoplable a MS. La separación HILIC despierta actualmente mucho interés ya que solventa muchos problemas previos de separación para ácidos orgánicos pequeños, drogas básicas y otras muchas sustancias polares neutras o cargadas. De hecho, se ha utilizado satisfactoriamente en el análisis de carbohidratos (Alpert, Shukla et al. 1994, Doco, Williams et al. 2015, Ghfar, Wabaidur et al. 2015), péptidos (Yoshida 2004, Hao, Lu et al. 2007), fármacos polares (Strege, Stevenson et al. 2000, Li and Huang 2004), etc.

Es importante destacar la importancia del uso de buffers o sales en esta técnica. En este caso, los tampones salinos, no solo facilitan la conservación del analito en una única forma iónica (o neutra), sino que pueden llegar a mejorar notablemente la eficiencia del proceso al aumentar el contenido salino de la ya mencionada capa de agua superficial. Este contenido salino permite un aumento de la polaridad de la fase estacionaria, pudiendo incrementar la retención. Es frecuente encontrar concentraciones de buffers, o acido/base, del orden de 5-10 mM o incluso mayores, mientras que en RPLC, suelen ser suficientes concentraciones 0.5-5 mM o 0.01-0.1% (w/v). Sin embargo, el uso de concentraciones elevadas de buffers, aunque se utilicen sales y/o ácidos volátiles, representa un inconveniente en el acoplamiento LC-MS pues, como se comentó anteriormente, la continua inyección de sales en la interfase de ionización afecta al rendimiento de ésta en análisis largos debido a la precipitación en el cono de extracción.

Las fases estacionarias utilizadas en las columnas HILIC pueden ser muy variables, siempre bajo la premisa de contener ligados grupos polares tales como dioles, ciano o amino. Al igual que en las columnas RPLC, la base de la fase estacionaria suele ser sílice o polimérica, aunque la primera no requiere derivatization pues los grupos silanol pueden actuar como fase estacionaria en este tipo de mecanismos.
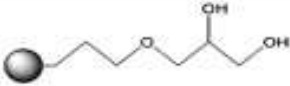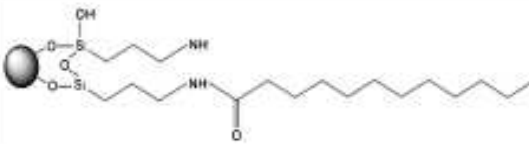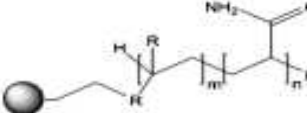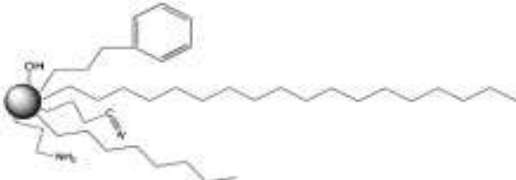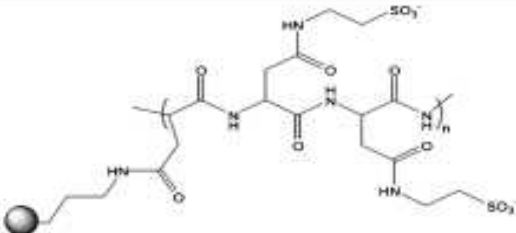
Packing materials	Structure of stationary phase
underivatized silica stationary phases that contain functional groups such as siloxanes, silanols with (or without) a small quantity of metals	
DIOL bonded phases	
cyano bonded phases	
amino bonded phases	
alkylamide	
amide bonded phases	
mix-mode	
polymeric structures of poly(succinimide) derivatives	

Figura 1.C. Algunos ejemplos de fase estacionarias utilizadas en HILIC

Cromatografía Liquida de Ultra Resolución (UHPLC)

Uno de los parámetros usados para evaluar la resolución cromatográfica de una columna es el número de platos teóricos, el cual es directamente proporcional a la longitud de la columna e inversamente proporcional a la altura equivalente a un plato teórico (HEPT, height equivalent to a theoretical plate). Para mejorar la eficacia de una columna se puede optar por:

a) una mayor longitud de la columna. Esto conlleva, además de tiempos de análisis más largos, un ensanchamiento de los picos por efectos de difusión, con la correspondiente perdida de sensibilidad (menor S/N).

b) una menor HEPT. La HEPT depende, entre otros factores de: la velocidad de flujo de la fase móvil, la homogeneidad de la fase estacionaria y el tamaño de partícula de ésta.

De acuerdo con la ecuación de Van Deemter para diferentes tamaños de partícula (figura 1.D.), se deducen fácilmente las ventajas que conlleva la reducción del tamaño de partícula, con la única limitación del aumento de la presión.

En los últimos años, se han desarrollado bombas de flujo capaces de trabajar eficiente y constantemente a las presiones requeridas para columnas de aproximadamente 1.7 µm. Esto ha implicado un notable avance en la cromatografía liquida, pasando a denominar como UHPLC (en lugar del común HPLC) a aquellas aplicaciones que trabajan con tamaños de partícula inferiores a las 2 µm. La UHPLC (Ultra-High Performance (or Pressure) Liquid Chromatography) no solo implica un notorio aumento de la resolución cromatográfica por la disminución de la HEPT, sino que además dicha disminución de HEPT es relativamente estable (grafica de Van Deemter más constante), con un mínimo a mayores flujos de la fase móvil. Esto ha permitido análisis mucho más rápidos y con una mejor separación cromatográfica, así como una mayor sensibilidad instrumental como consecuencia del estrechamiento de los picos.

Como se ha comentado anteriormente, la principal limitación para la implantación de UHPLC es la elevada presión generada por el empaquetamiento del

relleno de columna. Sin embargo, en lo que respecta al acoplamiento LC-MS, y más concretamente a la interfase ESI, existe un problema añadido pues los flujos de trabajo idóneos para esta interfase son normalmente inferiores a 0.3 mL/min. Por tanto, aunque la mejora en la resolución es notoria, hasta hace apenas un par de años no era posible maximizar las virtudes de los sistemas UHPLC acoplados a ESI. Es decir, no podían conseguirse la disminución de los tiempos de análisis y la máxima resolución óptimas dados los bajos flujos de trabajo requeridos, situándose en una situación intermedia entre HPLC y UHPLC (Figura 1.D.). Muy recientemente, se han desarrollado interfases ESI capaces de trabajar a flujos de fase móviles más elevados y adecuados para la UHPLC, ofreciendo buenos resultados a flujos de hasta 0.6-0.8 mL/min, gracias al aumento de la eficiencia en la desolvatación. Dicha eficiencia suele lograrse con mayores caudales de gas de desolvatación así como mayores temperaturas de dicho gas.
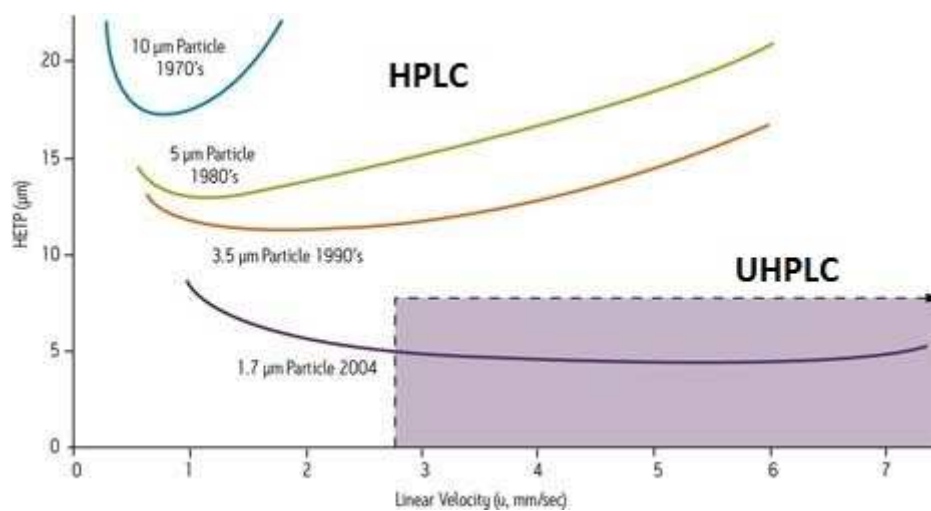
Figura 1.D. Representación de la HETP frente a la velocidad lineal de la fase móvil según la ecuación de Van Deemter para distintos tamaños de partícula de soporte de la fase estacionaria.

Esta nueva instrumentación ha posibilitado análisis en tiempos mucho más cortos y, consecuentemente, más económicos en prácticamente todas las áreas de conocimiento, desde análisis medioambientales, de alimentos, clínicos, etc. hasta el campo de la metabolómica.

1.1.2. Interfases LC-MS

Es un error bastante común pensar que los espectrómetros de masas son detectores universales, pues lo que se mide es la masa de los compuestos, una propiedad intrínseca de todas las moléculas. Sin embargo, es importante recordar que lo que en realidad se mide es la relación masa-carga (m/z) de los iones en fase gaseosa. Esto implica que la fuente o interfase de ionización juega uno de los papeles más cruciales en el acoplamiento LC-MS o GC-MS. De hecho, dicha fuente de ionización es más crucial si cabe en el acoplamiento LC-MS pues la GC presenta la ventaja de que los compuestos acceden a la fuente de ionización en forma gaseosa, facilitando en gran medida su ionización.
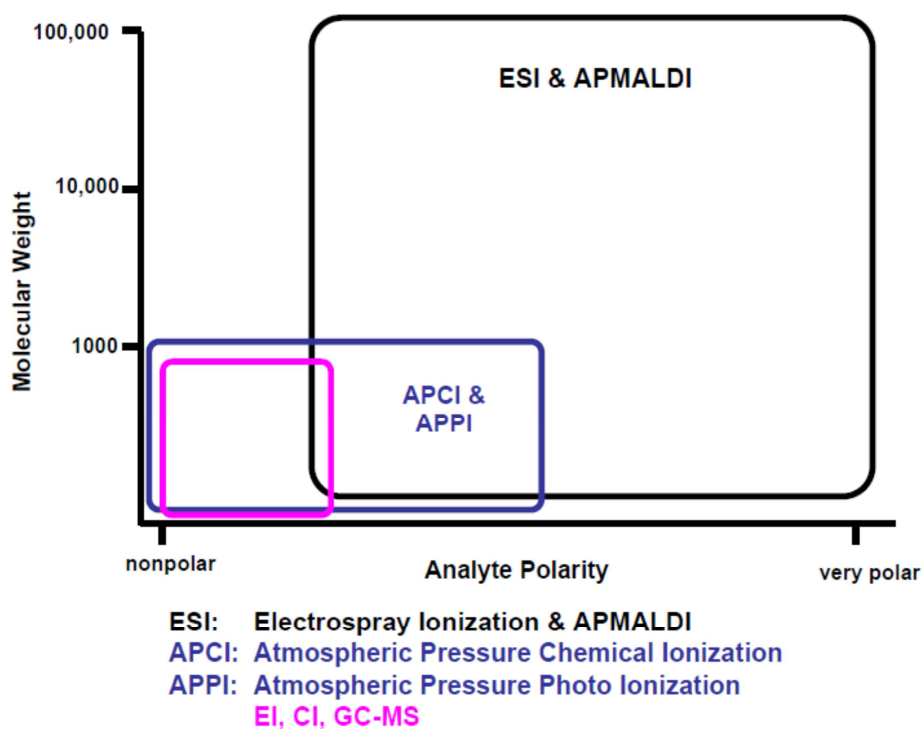
Figura 1.E. Comparación de la aplicabilidad de las diferentes API en función de la polaridad y el peso molecular de los analitos.

Si nos centramos en las fuentes de ionización o interfases actuales, es decir las API, podemos distinguir principalmente tres: ESI, APCI y APPI. Dichas interfases presentan diferentes ventajas y desventajas derivadas de su modo de desolvatación e ionización. La elección de una u otra interfase depende de la naturaleza del analito (Figura 1.E.), el tipo de matriz y del modo de separación aplicado (caudal de fase móvil, dimensiones de la columna, tipo de solventes utilizados, etc.).

A continuación se desarrollarán los aspectos prácticos más importantes a tener en cuenta para el uso de las diferentes API. En esta tesis doctoral se ha utilizado únicamente ESI, pero es importante distinguir las diferentes opciones para entender el porqué de dicha elección.

1.1.2.1. Electrospray ionization (ESI)

La interfase de <u>ionización por electrospray (ESI)</u> es sin lugar a duda la fuente de ionización más utilizada en el acoplamiento LC-MS. Esta se basa en la ionización del eluato proveniente de la LC mediante la aplicación de un alto voltaje (del orden de 3-5 kV) en un capilar de acero inoxidable, a través del cual son conducidos los analitos disueltos en la fase móvil antes de pasar al analizador de masas. Dicho voltaje produce la ionización de los analitos, así como de la fase móvil y la matriz. Mediante un flujo de gas inerte coaxial al capilar, normalmente N_2 a elevada temperatura (a unos 350 ºC), se produce la nebulización y desolvatación del líquido, formando un aerosol de pequeñas gotas cargadas que se hacen cada vez más pequeñas hasta que las fuerzas de repulsión entre cargas en la superficie son capaces de superar las fuerzas cohesivas de tensión superficial y se produce la ruptura final de la gota, originando iones en fase gaseosa (explosión de Coulomb).

La interfase ESI presenta flujos óptimos del orden de los pocos μL/min y está especialmente indicada para analitos polares, proporcionado la molécula protonada $[M+H]^+$, desprotonada $[M-H]^-$ o aductos con moléculas de solvente. Sin embargo, como se comentó en el apartado anterior, recientemente las interfases ESI han conseguido ser adaptadas a flujos mayores con la finalidad de hacerlas más compatibles con los flujos óptimos de técnicas UHPLC. Este desfase se ha resuelto

principalmente mediante el aumento del flujo y temperatura del gas de desolvatación. Así pues, las interfases ESI más recientes permiten trabajar a flujos de fase móvil cercanos a 1mL/min con rendimientos del proceso de ionización similares a los obtenidos a bajos flujos
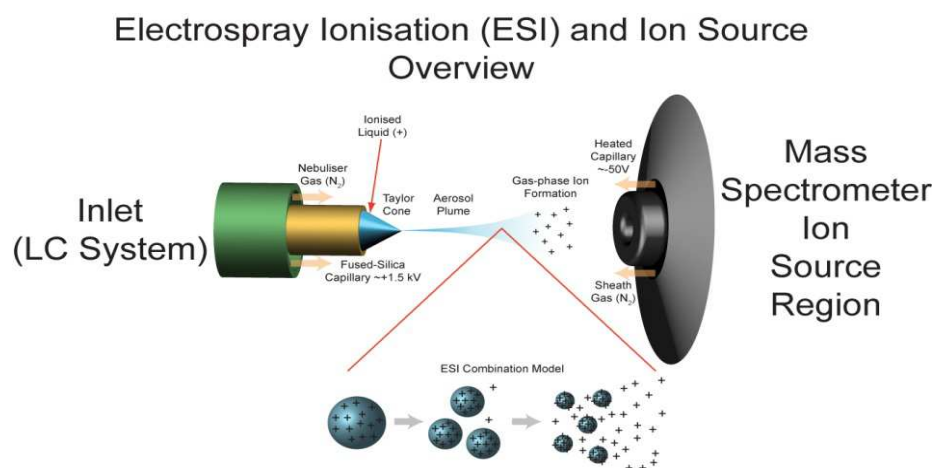
Figura 1.F. Representación esquemática del funcionamiento de la interfase ESI.

Por otra parte, la formación de aductos, especialmente aquellos formados con iones metálicos, no suele ser deseable porque es poco reproducible y producen una disminución de la sensibilidad. Además, en ciertos casos, como los aductos sodiados, la fragmentación suele ser pobre o requerir altas energías, lo cual complica la adquisición de fragmentos para confirmar la identidad del analito en cuestión. Sin embargo, para moléculas sin ningún centro ácido o básico se suelen añadir modificadores en la fase móvil o en el vial para facilitar su formación, siendo los más comunes los aductos sodiados y amoniados. Otra ventaja de ESI es que permite generar iones policargados, lo que amplía su rango de aplicación a compuestos de elevado peso molecular como péptidos, oligonucleótidos y oligosacáridos. Por este motivo, ha sido ampliamente utilizada en estudios bioquímicos y biomédicos. No obstante, la interfase ESI presenta también una serie de limitaciones. Por un lado, proporciona poca información estructural, ya que al tratarse de un sistema de ionización muy suave genera pocos fragmentos. Asimismo, la sensibilidad es muy

dependiente del flujo de fase móvil y del pH de ésta. Por último, la ionización mediante ESI es muy susceptible a sufrir efecto matriz, lo que provoca que se produzcan inhibiciones o exaltaciones indeseables de la señal (Marín, Gracia-Lor et al. 2009, Rossmann, Gurke et al. 2015).

1.1.2.2 APCI y APPI

Puesto que únicamente se ha utilizado la interfase ESI durante el desarrollo de esta tesis doctoral, este apartado se centrará exclusivamente en las diferencias más importantes entre las distintas interfases mencionadas.

Ambas técnicas funcionan de manera muy similar, pues los analitos son ionizados después del proceso de nebulización, es decir, en fase gaseosa. El proceso de ionización se realiza mediante una descarga de entre 2 y 5 kV proporcionada por una aguja corona (en el caso de la APCI) o mediante fotoionización con una lámpara UV (en la APPI).

Tabla 1.A. Principales diferencias entre la interfase ESI y APCI/APPI.

	ESI	**APCI/APPI**
Flujo	nL/min - µL/min	mL/min
Analitos	Polaridad alta o iónicos.	Polaridad media-baja.
	Grupos acido/básico u ionizables	No necesaria la presencia de grupos funcionales
	Termolábiles y termoestables	NO termolábiles
Iones	Capacidad de generar iones **policargados y aductos**	Iones **monocargados** y menor tendencia a formar de aductos
Respuesta	En función de la **concentración**	En función de la **cantidad**
Efecto matriz	Más influenciada	Menos influenciada
Fase móvil	Afectada por el pH	Afectada por el solvente orgánico

La interfase APPI requiere, en aquellos casos en los que los analitos presenten su primer potencial de ionización a mayor energía que la de los fotones emitidos por la lámpara UV, la adición de un agente dopante (por ejemplo, acetona, tolueno o anisol) que se ionice inicialmente para posteriormente transmitir la carga a los analitos. Por razones obvias, la interfase APPI es la menos ventajosa para afrontar un método de screening.

Por otro lado, basándonos en las características de los analitos susceptibles de ser analizados, se podría considerar la APCI como una opción intermedia entre la LC-ESI-MS y la GC-MS (Figura 1.E.). La elección de la interfase utilizada queda perfectamente justificada dado que el rango de compuestos analizables mediante ESI es mayor y más favorable para la aplicación que se pretende realizar en este caso.

1.1.3. Importancia de la resolución y la exactitud de masa

Si bien es cierto que el analizador de masas es probablemente la parte más distintiva y característica del instrumento, en esta introducción no se entrará a detallar los diferentes tipos y modos de acción pues sus funciones y limitaciones suelen ser sobradamente conocidas.

Para afrontar un reto tipo screening o un análisis no-dirigido como en el caso de las técnicas metabolómicas, la metodología debe cumplir, en mayor o menor medida, tres requisitos básicos: amplio rango de aplicación, sensibilidad y elevada información estructural.

La importancia del amplio rango de aplicación ya se trató al inicio de este capítulo. Obviamente, la amplitud de la metodología no depende exclusivamente del analizador utilizado. Seguramente, el analizador no sea el aspecto más influyente en la versatilidad analítica pero si puede contribuir a minimizar las limitaciones de otras partes del proceso.

En teoría todos los analizadores de masa pueden cubrir las necesidades espectrales para este tipo de retos ya que se trata en general de moléculas pequeñas

(generalmente con masas menores a 1000 uma). Sin embargo, existen dos tipos de modos de barrido cuando se trabaja en full-scan: secuencial o simultaneo. En el caso de aquellos instrumentos que por su propia naturaleza solo pueden realizar un barrido completo de forma secuencial, como es el caso de los cuadrupolos o el sector magnético, el rango de aplicación penaliza directamente la sensibilidad. Por tanto, la adquisición de un mayor rango de masas, repercutirá directamente en la intensidad de la señal. Se debe además tener en cuenta que, como se comentó en el apartado anterior, las técnicas UHPLC han permitido reducir los tiempos de análisis mejorando incluso la resolución cromatográfica, lo cual implica picos muy estrechos (del orden de los pocos segundos), lo que nos lleva a la necesidad de analizadores capaces de lidiar con estas velocidades de análisis. Por lo general, esta velocidad de análisis es incompatible con analizadores de masa que realicen un barrido completo de forma secuencial, al menos con una sensibilidad satisfactoria. Asimismo, resulta extremadamente complicado, en este tipo de instrumentos, la adquisición de un número suficiente de barridos durante la elución de un compuesto (típicamente se requieren más de 10 puntos por pico), aumentando así el error en la medida de la intensidad.

Si nos centramos pues en aquellos analizadores que, por su fundamento, realizan el escaneo de iones de forma simultánea encontramos diversas opciones, entre ellas la trampa de iones (y la trampa de iones cuadrupolar), TOF, Orbitrap y Fourier Transform Ion Cyclotron Resonance. Todos ellos ofrecen la posibilidad de evitar las desventajas del barrido secuencial. En realidad, la trampa de iones trabaja de forma secuencial, desestabilizando la trayectoria de los iones en función de su m/z, pero se incluye en el grupo de analizadores de barrido simultáneo pues éste es su modo de operación más común y su velocidad de barrido es realtivamente alta.

Si nos centramos pues en estos últimos analizadores, aquellos que realizan un barrido completo de iones, observamos que la mayor diferencia entre ellos es su resolución de masa.

El poder de resolución se define como la capacidad de un analizador de masas de distinguir entre dos iones de m/z muy cercana. Se trata pues de la propiedad más relevante en cuanto a información estructural se refiere. Para afrontar

cualquier análisis de tipo screening, resulta imprescindible poder confirmar la identidad del analito detectado e incluso poder elucidar la estructura química de un compuesto en cuestión. Para ello, un elevado poder de resolución permite discernir entre iones de masa muy cercana, facilitando enormemente la labor de identificación y elucidación.

El segundo aspecto fundamental en cuanto a información estructural se refiere es la exactitud de masa. Conviene recordar que los términos resolución y exactitud de masas, aunque suelen ir íntimamente ligados e incluso en algunos casos llegan a ser confundidos, se refieren a propiedades distintas. Mientras que el primero, como se explicó anteriormente, mide la capacidad de distinguir dos iones de masa muy próxima, el segundo se define como la capacidad instrumental de ofrecer un valor de m/z muy cercano al valor real. Si bien es cierto que el poder de resolución es condición *sine qua non* para la exactitud de masa robusta, existen otros factores necesarios o importantes que deben cumplirse como por ejemplo una correcta calibración del eje de masas. De hecho, la importancia de la correcta calibración del eje de masas se investiga en el primer artículo científico de la presente tesis.

La exactitud de masa, junto con el poder de resolución, permiten poder llevar a cabo tareas de elucidación, confirmación e identificación con una mayor probabilidad de éxito. En lo que respecta a la elucidación, y dado que el primer paso a realizar se trata normalmente de la obtención de la formula molecular del analito en cuestión, la exactitud de masa facilita enormemente esta labor ya que permite filtrar las posibles fórmulas elementales en un menor rango de masas, ofreciendo un mucho menor número de posibles candidatos (o incluso uno solo). Reduciendo el número de candidatos tanto para el ion pseudomolecular como en fragmentos (si los hubiera) se permite elucidar el analito en cuestión con un más elevado índice de acierto.

Pero no solo la capacidad de elucidación se ve mejorada considerablemente con la exactitud de masa. El poder de confirmación es, por razones obvias, más elevado. También la capacidad de detección e identificación se ve aumentada, pues el número de compuestos (ya sea endógenos o exógenos) que presenten una relación m/z idéntica o muy similar, y que por tanto, puedan ser detectados o identificados como el analito propiamente dicho, será menor. Además, puesto que para poder detectar cierto/s analito/s se requiere la extracción de sus narrow-window

XICs (nw-XICs), como se verá en más detalle posteriormente, también la S/N se ve notablemente mejorada ya que con ventanas de masas estrechas, las cuáles son posibles gracias a la resolución y exactitud de masas, el ruido espectral queda muy reducido e incluso eliminado en algunos casos.

Existen diversos analizadores que ofrecen un gran poder de resolución y una óptima exactitud de masa. Entre ellos encontramos los analizadores TOF, con un poder de resolución entre 10.000 y 60.000 FWHM (en el caso de los modelos más recientes), Orbitrap, entre 60.000 y 280.000 FWHM y FT-ICR, cuyo poder de resolución puede alcanzar el millón. Obviamente todos ellos presentan diferentes ventajas y desventajas como el coste (mayor al aumentar el poder de resolución), el rango de masas (mayor en TOF y menor en Orbitrap) y velocidad de barrido (Bijlsma, Emke et al. 2013). Cabe destacar acerca de esta última característica que, aunque el poder de resolución del Orbitrap es notoriamente mayor al de los analizadores TOF, éste depende inversamente de la velocidad de adquisición a la que se trabaje y, por tanto, a elevadas velocidades como las típicamente requeridas en el acoplamiento UHPLC-MS, el poder de resolución de ambos instrumentos es bastante similar.

1.1.4. Instrumentación híbrida. Modos de análisis

La espectrometría de masas en tándem (MS/MS) se puede llevar a cabo en el espacio, usando varios analizadores, o en el tiempo, mediante un único analizador que permite la fragmentación del ión atrapado en su interior y la posterior separación de los fragmentos producidos. Si nos centramos en los analizadores de alta resolución, casi todos ellos pertenecerían al primer grupo pues requieren de un analizador para preseleccionar la m/z deseada y una celda de colisión. Las posibles combinaciones dependen en gran medida de las aplicaciones deseadas y del presupuesto disponible, pudiendo encontrar desde el acoplamiento más sencillo QqTOF (QTOF) hasta el QqICRMS. Recientemente, se ha presentado un modelo de Orbitrap que podemos clasificar como tríbrido, ya que combina un cuadrupolo, una trampa de iones y un orbitrap, el Orbitrap Fusion™.

Sin embargo, el acoplamiento instrumental clásico MS/MS, entendido como análisis de iones productos, no supone una mejora real de la instrumentación hibrida de alta resolución a la hora de enfrentar retos como el que nos ocupa en esta tesis. Si bien es cierto que la información estructural del analito es muchísimo mayor, lo cual es realmente importante para la elucidación de compuestos o incluso para su correcta identificación, este modo de análisis es incompatible con aproximaciones tipo screening o metabolómicas. Por un lado, si se pre-seleccionan los iones en el primero de los analizadores (necesario en este modo de operación) la capacidad de barrido completo queda descartada y, por tanto, la capacidad de un screening no-dirigido, obteniéndose una metodología multiresidual con mayor poder de identificación y confirmación, comparado con instrumentos híbridos como QqQ o trampa de iones, pero con menor sensibilidad. La alternativa seria la realización de un análisis posterior al análisis no-dirigido para tratar de identificar aquellos analitos de interés. Obviamente esta segunda aproximación es tediosa y costosa por lo que se suele aplicar para casos estrictamente necesarios.

Afortunadamente existen alternativas que pueden sortear las limitaciones presentadas anteriormente. Dichas opciones están basadas en la adquisición de dos funciones, una que llamaremos de baja energía (LE) y otra de alta energía (HE). Para ello, se puede hacer uso tanto de la interfase de ionización (la forma más clásica) o, al igual en el modo MS/MS, de la celda de colisión (Hernández, Bijlsma et al. 2011). Esta última, requiere un instrumento hibrido que trabaje con el gas de colisión siempre presente en la celda. Esta aproximación, conocida como MS^E, permite la adquisición de dos funciones:

a) Función de baja energía (LE). En esta función no se aplica ningún tipo de energía de aceleración extra a los iones. Así pues, tanto si se aplica en la interfase como si se hace en la celda de colisión la fragmentación es mínima o inexistente, obteniéndose principalmente el ion (des)protonado u otros aductos dependiendo del analito en cuestión y de las condiciones de ionización. Como se ha comentado en el párrafo previo, si se utiliza un instrumento hibrido (como es el caso que aplica en esta tesis doctoral, contrariamente a lo que históricamente se ha realizado, la celda de colisión debe estar ocupada por gas ya que el proceso de llenado y vaciado de la

celda de colisión no es lo suficientemente rápido como para poder adquirir ambas funciones. Por tanto, aun tratándose de una función de MS sencilla (la LE), la celda de colisión resultará con presión de gas, de forma similar al modo de trabajo MS/MS. Sin embargo, al aplicar una energía de aceleración relativamente baja en la celda de colisión, los analitos no experimentan apenas reacciones de fragmentación.

b) Función de alta energía (HE). Alternativamente a la función LE se adquiere la HE que permite fragmentar todos los analitos que han sido ionizados, obteniendo un espectro más rico desde el punto de vista estructural. Este proceso se realiza mediante la aplicación de un voltaje elevado en el cono de extracción o en la celda de colisión, lo cual otorga a los iones la energía típicamente necesaria para poder obtener espectros de fragmentación. En ambos casos, no se realiza ningún tipo de preselección de los iones y, por tanto, no pueden considerarse como métodos MS/MS, si bien es cierto que, si se obtienen espectros libres de interferencias, éstos son muy similares a los obtenidos con MS/MS e incluso con mayor información al registrarse además los isotopos de cada uno de los fragmentos.

En el primer artículo científico de esta tesis se tratarán con más detalle las ventajas e inconvenientes de ambos modos de adquisición: MS/MS y MS^E.

1.2. ANALISIS TARGET Y NON-TARGET

Éste es sin duda el tema principal del segundo capítulo de esta tesis doctoral y más concretamente del segundo artículo científico. En el apartado anterior se han resumido las diferentes posibilidades instrumentales para afrontar aproximacions tipo screening "universal" y/o metabolómicas. Sin embargo, las posibilidades analíticas van más allá de lo puramente instrumental o químico. Resulta muy importante definir una estrategia para poder llevar a cabo dichas aproximaciones (Krauss, Singer et al. 2010). A continuación, se presentan algunas definiciones de conceptos clave aplicados durante toda la tesis doctoral y sus repercusiones en los diferentes campos estudiados.

1.2.1. Definiciones: análisis pre-target, post-target y non-target

Análisis pre-target: En este tipo de análisis, los compuestos a estudiar son previamente definidos y las condiciones de medida son específicas paras estos, por lo que cualquier análisis retrospectivo no es posible (Hernández, Ibáñez et al. 2014). Por definición, cualquier análisis en modo MS/MS, sin importar los analizadores que se utilizan, es un análisis pre-target puesto que requiere la preselección de los iones que se van a medir. Este tipo de análisis se lleva a cabo en métodos para compuestos individuales o multiresiduales con la finalidad de detectar, identificar y, generalmente, también cuantificar un cierto número de analitos target previamente seleccionados.

Las principales ventajas de este tipo de análisis son:

a)   Adecuación del método de análisis. Puesto que se trata de un número restringido de analitos conocidos, todo el proceso analítico desde la toma de muestra hasta el análisis instrumental puede ser optimizado adecuadamente con patrones de referencia.

b)   Mayor sensibilidad. La adquisición de un número limitado de iones y/o transiciones así como la optimización previa de las condiciones permiten que estos métodos multiresiduales sean, en lo general, más sensibles.

Se debe tener en cuenta además que, dado el conocimiento previo de la matriz y los analitos a analizar, la aplicación de métodos específicos de clean-up y pre-concentración puede incrementar la sensibilidad del método, llegando a niveles de concentración del orden de los sub-ppb, o incluso menores.

c)   Facilidad de procesamiento de datos. Dado que se adquiere un número limitado de datos, el tratamiento de esta información resulta más sencillo y rápido.

d)   Mayor capacidad de automatización y análisis de rutina. Este tipo de metodologías resultan mucho más convenientes a la hora de realizar análisis rutinarios de un gran número de muestras puesto que, además de facilitar el procesamiento de los datos, ofrecen una mayor versatilidad para la automatización de todo el proceso.

El principal inconveniente de este modo de análisis reside precisamente en la limitación a aquellos compuestos incluidos en el método, no permitiendo obtener información de otros posibles compuestos de interés, o sub-productos de aquellos analizados, ni ningún tipo de análisis retrospectivo. El análisis retrospectivo es el que se realiza a partir de los datos adquiridos en un análisis anterior, con el fin de investigar nuevos compuestos que no se consideraron inicialmente. Este tipo de análisis, aunque menos frecuente en el ámbito medioambiental, resulta extremadamente útil en campos como el anti-doping o salud publica, ya que frecuentemente se hace necesaria una re-evaluación de los resultados en base a nuevos datos o sospechas.

Las alternativas al análisis pre-target pasan por la adquisición en modo *full-scan* o barrido de iones. Este tipo de adquisición no requiere de ningún tipo de filtro o pre-selección de m/z, evitando así la limitación de compuestos a analizar. El hecho de utilizar este tipo de métodos no implica la universalidad del análisis pues existen muchos otros factores limitantes (como se comentará en el siguiente apartado) durante todo el proceso analítico. Sin embargo, las metodologías basadas en métodos *full-scan* presentan una mayor complejidad debido a la gran cantidad de información

generada, que precisa de herramientas adecuadas para poder centrar el problema en aquello que es realmente interesante. Una buena analogía de esta problemática seria la frase popular "encontrar una aguja en un pajar" puesto que solo una pequeña fracción de datos corresponde a los compuestos de interés, normalmente escondidos entre una gran mayoría de compuestos irrelevantes desde el punto de vista del análisis realizado. Podemos distinguir dos tipos de aproximaciones a la hora de encarar esta problemática (el imán necesario para encontrar la aguja): post-target y non-target.

Análisis post-target: Aunque no se trata de un término de uso estandarizado, se refiere a aquel análisis en el cual la adquisición (en este caso, mediante acoplamiento cromatografía-espectrometría de masas) se realiza de forma totalmente no dirigida, mientras que el tratamiento de datos se lleva a cabo en forma de cribado específico. En otras palabras, se trata de metodologías en las cuales la selección de los compuestos a investigar se realiza posteriormente a la adquisición de los datos. Sobre la base de adqusiciones de espectro completo (con masa exacta además en métodos basados en HRMS), se puede investigar, en princpio, un número ilimitado de compuestos en las muestras Lógicamente, el hecho de disponer de patrón de referencia facilita mucho esta búsqueda, pero también puede realizarse en teoría sin que exista el patrón de referencia. El Término post-target fue propuesto por nuestro grupo de investigación en 2005 (Hernández, Pozo et al. 2005).

Análisis non-target o untargeted: Al igual que en análisis post-target, no se realiza ninguna pre-selección de los compuestos a analizar pero, mientras que en el anterior se realiza un cribado específico, es decir, se investiga la presencia de compuestos concretos, el análisis non-target busca obtener a priori toda la información de la muestra sin ningún tipo de selección previa. Cualquier pico cromatográfico es, en principio, investigado sin importar su naturaleza. Posteriormente, se debe centrar el estudio en tan sólo ciertos compuestos, aplicando filtros (ej. presencia de Cl, S, o cualquier otro elemento; picos con una determinada abundancia, etc) pues la investigación completa del espectro completo es una aproximación poco realista (Ibáñez, Sancho et al. 2008).

Ambas aproximaciones eluden las carencias de las metodologías pre-target, permitiendo el análisis retrospectivo de los resultados, pero enfocan de manera distinta el problema. Mientras que la aproximación post-target se basa en responder a la cuestión sobre la presencia de determinados compuestos, la untargeted busca compilar toda la información de la muestra para poder resolver la pregunta a posteriori. Un análisis post-target se lleva a cabo mediante la extracción de cromatogramas XICs para las m/z de interés sin ningún tipo de limitación en cuanto al número, contrariamente a la aproximación pre-target donde el número de compuestos analizados es dependiente de los límites instrumentales o del método. Por otro lado, la aproximación non-target busca, mediante algoritmos de deconvolución, encontrar los picos cromatográficos presentes en la muestra, para la posterior elucidación o identificación de aquellos que puedan ser relevantes de algún modo.

Los algoritmos de deconvolución en espectrometría de masas, o más concretamente en acoplamientos cromatografía-MS, aplican diferentes criterios para poder extraer picos cromatográficos pertenecientes a un solo analito así como sus espectros de masas libres de interferencias. Estos algoritmos son altamente dependientes de la relación entre la S/N de los compuestos y la de sus interferentes.

En ambos casos, juega un papel esencial la resolución, tanto de masa como cromatográfica, y la exactitud de masa. Una elevada exactitud, con errores de masa muy bajos, permite poder procesar los XIC con ventanas de masa muy estrechas (los llamados narrow window XIC, nw-XIC), lo quefacilita la detección de picos debido al bajo ruido químico existente. Las diferentes ventajas de ambas aproximaciones dependen en gran medida de la aplicación requerida. A continuación, se discuten las posibilidades que ofrecen ambas para las aplicaciones estudiadas en esta tesis.

1.2.2. Screening de contaminantes orgánicos

Como ya se ha comentado anteriormente, el screening de contaminantes orgánicos en muestras medioambientales necesita de herramientas capaces de ofrecer unas prestaciones acorde con el número de compuestos y sus niveles de

concentración en las muestras. Algo similar aplica también para análisis anti-doping o de salud pública. Cada vez es mayor el número de compuestos que deben ser monitorizados por lo que es conveniente el uso de técnicas no dirigidas que ofrecen, además, la posibilidad de revisión de resultados frente a otros compuestos sospechosos de ser positivos en las muestras como, por ejemplo, productos de degradación no conocidos anteriormente o nuevos agentes dopantes. Cuando se debe analizar un alto número de muestras, la aplicación de metodologías non-target resulta tediosa y muy complicada en comparación a las clásicas metodologías multiresiduales. Por este motivo, el enfoque post-target se presenta como el más adecuado ya que los tiempos de procesamiento y revisión de datos se reducen notablemente con respecto a la metodología non-target, aunque se puede perder parte de la información contenida en la muestra al enfocar los análisis en compuestos concretos, aunque en número muy elevado. Podría verse la aproximación post-target como un compromiso entre los restringidos métodos pre-target y el amplio y complejo proceso untargeted (Ibáñez, Sancho et al. 2014).

La estrategia a seguir se inicia con la creación de una base de datos de compuestos de interés con información espectral y cromatográfica (si se posee). La información espectral abarcaría la m/z teórica del ion molecular (o pseudomolecular), así como posibles aductos que pudieran generarse y, si está disponible, m/z de los fragmentos más importantes. Posteriormente, gracias a la adquisición en modo full-scan y a la exactitud de masa usando técnicas de HRMS, se procedería a la detección de los compuestos incluidos en la base de datos. Cuando se dispone de suficiente información, se pueden identificar tentativamente los analitos encontrados y confirmar dicha identidad si se dispone del patrón de referencia. Incluso se puede realizar una estimación de sus niveles de concentración si se han inyectado los patrones correspondientes.

La aproximación non-target conlleva una metodología de actuación completamente diferente. En este caso, el primer paso a seguir es la búsqueda de todos los diferentes compuestos presentes que dan lugar a picos en el cromatograma, para posteriormente intentar identificar, entre los compuestos encontrados, posibles

analitos de interés. Por razones obvias, este proceso requiere de algún tipo de automatización que posibilite discernir dichos candidatos. Normalmente, este proceso se realiza mediante la comparación de los espectros deconvolucionados frente a una librería de espectros (Ibáñez, Sancho et al. 2008).

Ambas aproximaciones, así como las diferentes opciones dentro de ellas, se tratan con profundidad en el artículo científico 2 dentro del segundo capítulo de esta Tesis. En dicho artículo se lleva a cabo una extensa discusión sobre los pros y contras en función del objetivo que se pretende alcanzar para el screening de contaminantes orgánicos en diferentes muestras de origen ambiental, alimentario y biológico.

1.2.3. Metabolómica

Existen diferentes definiciones del término metabolómica pues se trata de una disciplina científica relativamente reciente que engloba una gran cantidad de estudios y procesos. Como el resto de –ómicas, como la genómica o la proteómica, la metabolómica estudia los diferentes fenómenos biológicos que ocurren en todo tipo de organismos. A diferencia de la genómica y la proteómica, que estudian la composición e interrelación del ADN (o más concretamente los genes) y las proteínas, respectivamente, la metabolómica se basa en la composición de los metabolitos presentes en diferentes tejidos y matrices. Más concretamente se pretende discernir o predecir cambios en la composición metabólica de un determinado organismo en base a un estímulo externo o cambios internos (Lindon, Holmes et al. 2003).

Dicha definición resulta extremadamente amplia pues abarca desde el estudio de enfermedades, con la finalidad de un diagnóstico temprano o de un mayor conocimiento de la misma, a la investigación de los efectos de un determinado medicamento o droga. La metabolómica se extiende incluso más allá de la investigación médica (Calderón-Santiago, Priego-Capote et al. 2015, Dickens, Larkin et al. 2015, McGrath and Young 2015) o farmacéutica (Puchades-Carrasco and Pineda-Lucena 2015, Vincent and Barrett 2015), pues es también aplicada en campos como la

seguridad alimentaria (German and Watkins 2004), en plantas (Teo, Tan et al. 2011, Ni, Yang et al. 2015, Zhao, Zhang et al. 2015) o el anti-doping (Raro, Ibáñez et al. 2015), entre otros.

También resulta amplia la definición en términos prácticos pues, el estudio global de los metabolitos presentes en una determinada muestra es un reto analítico de gran dificultad. Por ello resulta necesario en ocasiones centrar el estudio en un limitado número de compuestos en base a cierto conocimiento previo del problema. Se trataría de una metabolómica dirigida (target) denominada metabolite profilling (Fiehn 2002), la cual facilita enormemente el problema pero que, por otra parte, limita en gran medida el entendimiento del mismo. Se suele aplicar a metabolitos que juegan algún tipo de rol en determinadas rutas metabólicas que participan en el proceso en cuestión o a una familia de compuestos que pueda tener influencia, como por ejemplo lípidos (Bamba, Lee et al. 2012) o aminoácidos (Lee, Jung et al. 2015).

Sin embargo, la aproximación más común es el metabolic fingerprint o screening global y completamente no-dirigido con la finalidad de clasificar las muestras en base a patrones metabólicos o huellas dactilares (fingerprints) que cambian en función de un estímulo ambiental o como respuesta a una enfermedad o perturbación genética, con el fin último de identificar metabolitos discriminantes. En esta aproximación se realiza un análisis non-target de las muestras, mediante aplicación de los ya mencionados algoritmos de deconvolución o mecanismos similares, con el objetivo final de obtener la mayor información de la muestra posible. Posteriormente, aunque existen muchas bases de datos y librerías de espectros especializadas en diversos tipos de compuestos y matrices (como por ejemplo; NIST, MassBank o METLIN), es un requisito indispensable el uso de técnicas estadísticas para poder discernir previamente los metabolitos involucrados o significantes en el proceso metabólico o el estímulo en cuestión.

Las diferentes posibilidades tanto de screening como de procesamiento estadístico de datos se discuten en el correspondiente capítulo de esta tesis doctoral dedicado a la aplicación de técnicas metabolómicas para la discriminación de alimentos en función de su origen.

1.3. TRATAMIENTO DE MUESTRA

En los puntos anteriores se han resumido las ventajas y limitaciones de las diferentes técnicas analíticas aplicadas, así como las diferentes posibilidades en cuanto a la estrategia analítica seguida y al tratamiento de datos. Ciertamente, los acoplamientos cromatografía/HRMS resultan indispensables para afrontar análisis de tipo screening (como nos ocupa en esta Tesis). Sin embargo, muchas veces se resta importancia a los pasos previos del proceso analítico, los cuales resultan tanto o más importantes como el análisis instrumental o el tratamiento de datos. El planteamiento del problema afrontado, con la adecuada selección de la estrategia a seguir, incluyendo la elección de las técnicas utilizadas, la adecuada selección y toma de muestra y el tratamiento de dichas muestras a lo largo del proceso analítico resultan pasos cruciales para la consecución de los objetivos marcados.

La selección y la toma de muestras son pasos esenciales dentro del proceso analítico, a los cuales no se les suele dar la importancia que merecen, tal vez porque la dificultad que entrañan los análisis y el tratamiento de datos en los complejos análisis de residuos en el campo medioambiental y alimentario hace que la mayoría de químicos analíticos centren sus esfuerzos en esta última parte del proceso. De nada sirve una correcta cuantificación de la muestra si no se puede asegurar que dicha muestra es representativa de la población en estudio. Una ajustada selección de la cantidad de muestra requerida, la situación y las condiciones de la toma de muestra, o las condiciones de transporte y conservación son aspectos fundamentales para asegurar la validez de los resultados. Esto cobra más importancia, si cabe, en el caso de la metabolómica ya que, al tratarse de estudios totalmente no dirigidos en los que suele ser necesario el análisis de numerosas muestras, requieren que la toma de dichas muestras cumpla ciertos requisitos. Así pues, en este caso es necesario distribuir aleatoriamente todo el proceso analítico, evitando así posibles interferencias por errores sistemáticos no detectados. Además, es importante establecer correctamente el número y la distribución de muestras para que sean representativativas de la población y/o fijar aquellas variables que no sean objeto de estudio. Pongamos por

ejemplo que se quiere estudiar el efecto metabólico de un fármaco mediante el análisis de una cierta muestra biológica con una aproximación metabolómica. En este caso, variables como género, edad, tipo de alimentación (y un largo etcétera de posibilidades) deberán ser controladas, bien sea fijando determinadas variables en el objeto de estudio (por ejemplo, centrando el análisis únicamente en mujeres entre 30 y 40 años) o distribuyendo aleatoriamente la selección de los individuos y asegurando una cantidad representativa de los mismos.

En lo que a la conservación de las muestras se refiere, depende de la naturaleza de las mismas y de los analitos en cuestión. Probablemente, las dos alternativas más comunes son la congelación y la liofilización. En general, se suele asumir que ambas técnicas, especialmente el mantenimiento de las muestras congeladas, es condición suficiente para asegurar la estabilidad de las mismas. A pesar de ello, es recomendable la realización de estudios previos que permitan establecer las condiciones más favorables para la conservación previa a los análisis en el laboratorio.

Por otro lado, el tratamiento de muestra puede llegar a ser el cuello de botella del proceso analítico global. Resulta imprescindible tener en cuenta la globalidad del proceso analítico para llevar a cabo la aproximación más correcta en cada caso, dependiendo de la naturaleza de la muestra y los analitos, así como el análisis instrumental que se pretende realizar.

1.3.1. "Universalidad"

Cuando se pretende realizar un análisis no dirigido de las muestras, ya sea con el fin de llevar a cabo un *wide-scope screening* o una aplicación metabolómica, se requiere un tratamiento de muestra "universal". Por obvias razones, la universalidad resulta imposible (de ahí su entrecomillado), por lo que se debe encontrar una solución lo más amplia posible, adecuada para el tipo de muestra y posterior análisis, donde las pérdidas de analito sean las mínimas posibles.

Normalmente, la universalidad en el tratamiento de muestra va estrechamente ligada a su sencillez ya que a mayor número de procesos realizados, mayor es la probabilidad de perder analitos en algunas de las etapas del análisis. A continuación, se resumen algunas de las opciones disponibles para afrontar dicho tratamiento en función del estado de la muestra (Castro-Puyana and Herrero 2013, Arbulu, Sampedro et al. 2015).

A) MUESTRAS LÍQUIDAS

1) Inyección directa

Sin lugar a dudas, si la muestra lo permite, la inyección directa es el tratamiento más universal y por tanto más adecuado para este tipo de análisis. En realidad, en este caso no existe ningún tratamiento de muestra por lo que rigurosamente no debería estar enmarcado en este punto. La ausencia de cualquier tratamiento permite tener una visión global sin pérdida alguna de analitos en el proceso. Sin embargo, no siempre es posible realizar una inyección directa, bien sea por incompatibilidad de la muestra o inapropiada concentración de los analitos. La incompatibilidad de la muestra puede deberse al solvente que las constituye, a compuestos interferentes o a la complejidad en su composición, lo que hace que no puedan ser directamente inyectadas.

2) Dilute-and-shoot

La alternativa más común y simple a la inyección directa suele ser la dilución previa de la muestra. Mediante dicha dilución se reduce el efecto matriz, se puede eliminar o minimizar el efecto de interferentes o la suciedad de la muestra (por ejemplo, precipitación de proteínas mediante un disolvente orgánico), además de corregir incompatibilidades con el solvente y reducir la concentración de los analitos (en el caso de que ésta fuera demasiado elevada). Idealmente, este proceso puede ir acompañado de una centrifugación y/o filtración de la muestra con el fin de evitar

posibles solidos disueltos que interfieran o dificulten el posterior análisis. Sin embargo, la filtración no es siempre recomendada pues puede ser selectiva y retener compuestos de interés en el filtro.

Esta simple aproximación también permite minimizar problemas de deriva de la señal instrumental en secuencias muy largas, las cuales sufren a menudo una pérdida progresiva de la respuesta debido a la deriva instrumental.

3) SPE

La extracción en fase solida (SPE) es un proceso selectivo y, por tanto, debe evitarse en casos en los que la "universalidad" es la característica a conseguir. Sin embargo, en el caso del análisis de contaminantes orgánicos, y más concretamente en muestra ambientales, resulta casi obligado realizar una pre-concentración de los analitos puesto que estos suelen estar a niveles de trazas.

Existe una gran variedad de fases disponibles para llevar a cabo la extracción pero se deben buscar aquellas con un espectro de aplicación lo más amplio posible para retos del tipo *wide-scope screening*. Dependiendo del tipo de análisis posterior existen diversas opciones como bases poliméricas con mecanismo de retención en fase reversa, cartuchos con fase estacionaria alquílica (por ejemplo, C18) o incluso cartuchos mixtos combinando la retención en fase reversa y el intercambio iónico. La elección depende principalmente del analito y de los interferentes de la matriz así como de la técnica analítica utilizada a posteriori para el análisis.

Resulta clave la elección del pH óptimo de trabajo pues ambos, fase estacionaria y analitos, deben presentarse en la correcta forma para que la extracción sea exitosa, llegando a ser altamente aconsejable en algunos casos la realización de réplicas a distintos valores de pH, tanto ácidos como alcalinos. Por esta razón, es obligatorio conocer las limitaciones de un método analítico de este tipo al menos desde un punto de vista cualitativo (como se profundiza en el tercer artículo científico

de esta tesis) para poder establecer qué analitos son identificados y a qué niveles de concentración.

4) LLE, SPME, stir bar extraction

Aunque diferentes, las tres técnicas son una alternativa eficiente, rápida y relativamente universal para los casos en que se exista una incompatibilidad de solventes con la técnica analítica utilizada y/o se requiera cierta pre-concentración de la muestra. Normalmente, suelen ir asociadas al análisis de muestras acuosas mediante GC.

5) Evaporación, liofilización

Al igual que las anteriores técnicas, estas son altamente aconsejables con la finalidad de pre-concentrar la muestra y efectuar un cambio de disolvente. Especialmente la liofilización es muy utilizada en muestras biológicas pues, además de permitir la mejor conservación de la muestra hasta el momento del análisis, la sublimación del disolvente se realiza a bajas temperaturas (con la muestra congelada), dificultando posibles reacciones de degradación de los metabolitos y asegurando su mejor conservación.

B) MUESTRAS SÓLIDAS

1) Extracción con solvente (asistida)

Como se ha comentado anteriormente, el objetivo de este tipo de aproximaciones es buscar la sencillez y la "universalidad" del proceso de extracción. En el caso de muestras sólidas, la extracción sólido-líquido reúne, en general, dichos requisitos. Sin embargo, en el caso de aproximaciones metabólicas, se suele requerir de una extracción múltiple con el fin de poder abarcar distintos tipos de metabolitos.

Como norma general el factor decisivo suele ser la polaridad del solvente de extracción, realizándose múltiples extracciones secuenciales con solventes de polaridad decreciente. Es muy común la realización de una primera extracción con solvente acuoso o una mezcla acuosa con un solvente orgánico relativamente polar para extraer así los compuestos de mayor polaridad. Posteriormente, una extracción con solventes apolares permite ampliar el espectro de aplicación al obtener la fracción de metabolitos más apolar de la muestra, como por ejemplo, analitos liposolubles. Los diferentes extractos, analizados según el tipo de metabolitos en cada fracción, ofrecen una visión amplia y complementaria de la muestra.

Sin embargo, otros factores como el pH y la temperatura pueden también tener un efecto notable en el proceso de extracción. Lamentablemente, no siempre es posible poder abarcar todas las posibilidades y la toma de decisiones depende del conocimiento que se tenga de la dupla muestra-analito. Existen a su vez técnicas que permiten aumentar la eficiencia de la extracción o disminuir el tiempo del proceso (a modo de ejemplo la extracción asistida por microondas o por ultrasonidos). Aunque pueden resultar muy ventajosas, su utilización puede ser selectiva y son por tanto recomendables cuando se tiene la certeza y conocimiento suficiente de la validez para la muestra y analitos de interés.

2) Análisis directo

El análisis directo de muestras solidas es una alternativa óptima si se dispone de la instrumentación adecuada. Fuentes de ionización como MALDI o DART, entre otras, permiten el análisis directo de muestras sólidas. Como principal inconvenientes, a parte de la necesidad de disponer de dichos instrumentos, estarían la elevada dificultad del tratamiento de datos al no existir separación cromatográfica previa, la representatividad y homogenidad de la porción de muestas sólida sometida al análisis y que se trata de un análisis de la superficie.

3) SPME

La microextracción en fase solida es una excelente alternativa para el análisis de compuestos volátiles o semivolátiles en muestras sólidas. Se suele aplicar, en modo *Head Space*, cuando el análisis ser realiza mediante GC-MS debido a la facilidad de acoplamiento.

1.3.2. Limitaciones

Existen unos cuantos trabajos publicados en los cuales se llevan a cabo métodos screening de contaminantes en diferentes tipos de muestras. En general, todos ellos muestran las bondades de este tipo de aproximación pero no es frecuente encontrar una discusión crítica sobre las limitaciones de la metodología aplicada. En algunos casos, no se muestra ninguna optimización del proceso o estudio previo y se centran únicamente en los resultados de aquellos analitos identificados. Sin embargo, resulta de suma importancia conocer las limitaciones de la metodología aplicada con el fin de, al menos, evaluar qué tipo de analitos pueden ser correctamente detectados e identificados y en qué rango de concentraciones. En esta Tesis se ha llevado acabo una validación cualitativa, cuya finalidad es la de evaluar, para un numero de analitos considerable y tomados como modelo, la capacidad de detección e identificación de la metodología aplicada a diferentes rangos de concentraciones.

1.4. REFERENCIAS

Alpert, A. J., M. Shukla, A. K. Shukla, L. R. Zieske, S. W. Yuen, M. A. Ferguson, A. Mehlert, M. Pauly and R. Orlando (1994). "Hydrophilic-interaction chromatography of complex carbohydrates." <u>J Chromatogr A</u> **676**(1): 191-122.

Arbulu, M., M. C. Sampedro, A. Gómez-Caballero, M. A. Goicolea and R. J. Barrio (2015). "Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines." <u>Analytica Chimica Acta</u> **858**(1): 32-41.

Bamba, T., J. W. Lee, A. Matsubara and E. Fukusaki (2012). "Metabolic profiling of lipids by supercritical fluid chromatography/mass spectrometry." <u>Journal of Chromatography A</u> **1250**: 212-219.

Bijlsma, L., E. Emke, F. Hernández and P. De Voogt (2013). "Performance of the linear ion trap Orbitrap mass analyzer for qualitative and quantitative analysis of drugs of abuse and relevant metabolites in sewage water." <u>Analytica Chimica Acta</u> **768**(1): 102-110.

Calderón-Santiago, M., F. Priego-Capote, N. Turck, X. Robin, B. Jurado-Gámez, J. C. Sanchez and M. D. Luque De Castro (2015). "Human sweat metabolomics for lung cancer screening." <u>Analytical and Bioanalytical Chemistry</u> **407**(18): 5381-5392.

Castro-Puyana, M. and M. Herrero (2013). "Metabolomics approaches based on mass spectrometry for food safety, quality and traceability." <u>TrAC - Trends in Analytical Chemistry</u> **52**: 74-87.

Dickens, A. M., J. R. Larkin, B. G. Davis, J. L. Griffin, T. D. W. Claridge, N. R. Sibson and D. C. Anthony (2015). "NMR-Based Metabolomics Separates the Distinct Stages of Disease in a Chronic Relapsing Model of Multiple Sclerosis." <u>Journal of Neuroimmune Pharmacology</u> **10**(3): 435-444.

Doco, T., P. Williams, E. Meudec, V. Cheynier and N. Sommerer (2015). "Complex carbohydrates of red wine: Characterization of the extreme diversity of neutral oligosaccharides by ESI-MS." <u>Journal of Agricultural and Food Chemistry</u> **63**(2): 671-682.

Fiehn, O. (2002). "Metabolomics – the link between genotypes and phenotypes." <u>Plant Molecular Biology</u> **48**(1-2): 155-171.

German, J. B. and S. M. Watkins (2004). "Metabolic assessment—a key to nutritional strategies for health." <u>Trends in Food Science & Technology</u> **15**(11): 541-549.

Ghfar, A. A., S. M. Wabaidur, A. Y. B. H. Ahmed, Z. A. Alothman, M. R. Khan and N. H. Al-Shaalan (2015). "Simultaneous determination of monosaccharides and oligosaccharides in dates using liquid chromatography-electrospray ionization mass spectrometry." <u>Food Chemistry</u> **176**: 487-492.

Hao, Z., C. Y. Lu, B. Xiao, N. Weng, B. Parker, M. Knapp and C. T. Ho (2007). "Separation of amino acids, peptides and corresponding Amadori compounds on a silica column at elevated temperature." <u>J Chromatogr A</u> **1147**(2): 165-171.

Hernández, F., L. Bijlsma, J. V. Sancho, R. Díaz and M. Ibáñez (2011). "Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography-quadrupole-time-of-flight-mass spectrometry." <u>Analytica Chimica Acta</u> **684**(1-2): 96-106.

Hernández, F., M. Ibáñez, R. Bade, L. Bijlsma and J. V. Sancho (2014). "Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry." <u>TrAC - Trends in Analytical Chemistry</u> **63**: 140-157.

Hernández, F., M. Ibáñez, T. Portolés, M. I. Cervera, J. V. Sancho and F. J. López (2015). "Advancing towards universal screening for organic pollutants in waters." <u>Journal of Hazardous Materials</u> **282**: 86-95.

Hernández, F., O. J. Pozo, J. V. Sancho, F. J. López, J. M. Marín and M. Ibáñez (2005). "Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC-MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers." <u>TrAC - Trends in Analytical Chemistry</u> **24**(7): 596-612.

Ibáñez, M., J. V. Sancho, L. Bijlsma, A. L. N. Van Nuijs, A. Covaci and F. Hernández (2014). "Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances." <u>TrAC - Trends in Analytical Chemistry</u> **57**: 107-117.

Ibáñez, M., J. V. Sancho, F. Hernández, D. McMillan and R. Rao (2008). "Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-light mass spectrometry." <u>TrAC - Trends in Analytical Chemistry</u> **27**(5): 481-489.

Krauss, M., H. Singer and J. Hollender (2010). "LC-high resolution MS in environmental analysis: From target screening to the identification of unknowns." <u>Analytical and Bioanalytical Chemistry</u> **397**(3): 943-951.

Lee, J., Y. Jung, J. Y. Park, S.-H. Lee, D. H. Ryu and G.-S. Hwang (2015). "LC/MS-based polar metabolite profiling reveals gender differences in serum from patients with myocardial infarction." <u>Journal of Pharmaceutical and Biomedical Analysis</u> **115**: 475-486.

Li, R. and J. Huang (2004). "Chromatographic behavior of epirubicin and its analogues on high-purity silica in hydrophilic interaction chromatography." <u>J Chromatogr A</u> **1041**(1-2): 163-169.

Lindon, J. C., E. Holmes and J. K. Nicholson (2003). "So what's the deal with metabonomics?" <u>Anal Chem</u> **75**(17): 384a-391a.

Marín, J. M., E. Gracia-Lor, J. V. Sancho, F. J. López and F. Hernández (2009). "Application of ultra-high-pressure liquid chromatography-tandem mass spectrometry to the determination of multi-class pesticides in environmental and wastewater samples: Study of matrix effects." <u>Journal of Chromatography A</u> **1216**(9): 1410-1420.

McGrath, C. M. and S. P. Young (2015). "Lipid and Metabolic Changes in Rheumatoid Arthritis." <u>Current Rheumatology Reports</u> **17**(9).

Ni, J., X. Yang, J. Zhu, Z. Liu, Y. Ni, H. Wu, H. Zhang and T. Liu (2015). "Salinity-induced metabolic profile changes in Nitraria tangutorum Bobr. suspension cells." <u>Plant Cell, Tissue and Organ Culture</u> **122**(1): 239-248.

Poole, C. F. (2003). <u>The essence of chromatography</u>. Ed. Elsevier. 1st Edition.

Portolés, T., E. Pitarch, F. J. López and F. Hernández (2011). "Development and validation of a rapid and wide-scope qualitative screening method for detection and identification of organic pollutants in natural water and wastewater by gas chromatography time-of-flight mass spectrometry." <u>Journal of Chromatography A</u> **1218**(2): 303-315.

Puchades-Carrasco, L. and A. Pineda-Lucena (2015). "Metabolomics in pharmaceutical research and development." <u>Current Opinion in Biotechnology</u> **35**: 73-77.

Raro, M., M. Ibáñez, R. Gil, A. Fabregat, E. Tudela, K. Deventer, R. Ventura, J. Segura, J. Marcos, A. Kotronoulas, J. Joglar, M. Farré, S. Yang, Y. Xing, P. Van Eenoo, E. Pitarch, F. Hernández, J. V. Sancho and O. J. Pozo (2015). "Untargeted Metabolomics in Doping Control: Detection of New Markers of Testosterone Misuse by Ultrahigh Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry." <u>Analytical Chemistry</u> **87**(16): 8373-8380.

Richardson, S. D. and T. A. Ternes (2014). "Water analysis: Emerging contaminants and current issues." <u>Analytical Chemistry</u> **86**(6): 2813-2848.

Rossmann, J., R. Gurke, L. D. Renner, R. Oertel and W. Kirch (2015). "Evaluation of the matrix effect of different sample matrices for 33 pharmaceuticals by post-column infusion." <u>J Chromatogr B Analyt Technol Biomed Life Sci</u> **1000**: 84-94.

Snyder, L. R. K., J.J.; Dolan, J.W. (2009). <u>Introduction to modern liquid chromatography</u>. Ed. Willey. 3rd Edition.

Strege, M. A., S. Stevenson and S. M. Lawrence (2000). "Mixed-mode anion-cation exchange/hydrophilic interaction liquid chromatography-electrospray mass spectrometry as an alternative to reversed phase for small molecule drug discovery." <u>Anal Chem</u> **72**(19): 4629-4633.

Teo, C. C., S. N. Tan, J. W. Hong Yong, T. Ra, P. Liew and L. Ge (2011). "Metabolomics analysis of major metabolites in medicinal herbs." <u>Analytical Methods</u> **3**(12): 2898-2908.

Vincent, I. M. and M. P. Barrett (2015). "Metabolomic-based strategies for anti-parasite drug discovery." <u>Journal of Biomolecular Screening</u> **20**(1): 44-55.

Wishart, D. S. (2008). "Quantitative metabolomics using NMR." <u>TrAC - Trends in Analytical Chemistry</u> **27**(3): 228-237.

Yoshida, T. (2004). "Peptide separation by Hydrophilic-Interaction Chromatography: a review." <u>J Biochem Biophys Methods</u> **60**(3): 265-280.

Zhao, Y., L. Zhang, C. Zhao, C. Hu, Y. Li, J. Zhao, J. Zhang, L. Li, Y. Chang, F. Wang, X. Lu, Z. Zhu and G. Xu (2015). "Metabolic responses of rice leaves and seeds under transgenic backcross breeding and pesticide stress by pseudotargeted metabolomics." <u>Metabolomics</u>.

CAPÍTULO 2

SCREENING "UNIVERSAL" DE CONTAMINANTES ORGÁNICOS EN MUESTRAS DE INTERÉS AMBIENTAL, ALIMENTARIO Y TOXICOLÓGICO.

CAPÍTULO 2. SCREENING "UNIVERSAL" DE CONTAMINANTES ORGÁNICOS EN MUESTRAS DE INTERÉS AMBIENTAL, ALIMENTARIO Y TOXICOLÓGICO

2.1. INTRODUCCIÓN

El término contaminante orgánico abarca una ingente cantidad de compuestos, que tienen unas características, propiedades físico-químicas, usos y efectos muy diversos. Éstos se pueden agrupar según la familia a la que pertenecen (por ejemplo, pesticidas, fármacos, micotoxinas, etc.) y sus efectos pueden afectar desde distintos seres vivos hasta a humanos (Richardson 2009). El uso cada vez mayor de diferentes productos y los actuales hábitos de consumo, aunque se traduce en una mejor calidad de vida, también genera una gran cantidad de residuos químicos que se están convirtiendo en una amenaza potencial para la calidad del agua (aunque también en alimentos) y como consecuencia, para el medio ambiente y los seres vivos. Estos contaminantes, generalmente de origen antropogénico, se están introduciendo continuamente en el medioambiente y afectan de manera negativa a los ecosistemas. Aunque el control de estos compuestos es cada vez mayor, el continuo desarrollo de nuevos productos dificulta enormemente esta labor, requiriéndose una actualización periódica de las listas de contaminantes a controlar en todo tipo de muestras (Gracia-Lor, Sancho et al. 2010).

El término **"contaminantes emergentes"** se refiere a compuestos que no habían sido considerados peligrosos hasta ahora, o que eran desconocidos, aun cuando su presencia en el ambiente no sea necesariamente nueva. Son compuestos de distinto origen y naturaleza química, que no se encuentran todavía sometidos a regulación. Los contaminantes emergentes incluyen una amplia variedad de productos de uso diario con aplicaciones tanto industriales como domésticas: surfactantes, productos farmacéuticos, drogas de abuso, productos para el cuidado personal, aditivos de

gasolinas, retardantes de llama, antisépticos, aditivos industriales, esteroides, hormonas y subproductos de la desinfección del agua (Tabla 2.A.). Su presencia y posibles efectos sobre el ambiente han pasado muchas veces inadvertidos, por lo cual son necesarias investigaciones acerca de su impacto, su presencia en diferentes fuentes de agua y la evaluación de tecnologías para su degradación. Por ese motivo, han sido objeto frecuente de estudio en la última década, siendo numerosas las publicaciones relativas a contaminantes emergentes en los últimos años (Petrovic, Gonzalez et al. 2003).

Los "contaminantes emergentes" corresponden en la mayoría de los casos a compuestos no regulados, que pueden ser candidatos a regulación futura dependiendo de las investigaciones sobre sus efectos potenciales en la salud y de los datos sobre niveles de concentración en el medio ambiente. La principal característica de estos contaminantes es que no necesitan ser persistentes para causar efectos negativos, puesto que sus altas tasas de transformación/eliminación se pueden compensar por su introducción continua en el ambiente (Barceló 2003, Ellis 2006).

Aunque estos compuestos se encuentran en concentraciones muy bajas en el medio ambiente, constituyen un riesgo para la salud de la población humana y de otras especies animales (Richardson 2009). Los contaminantes emergentes no son necesariamente persistentes, pero debido a su solubilidad en agua son capaces de afectar a todas las etapas del ciclo del agua (Figura 2.A.), planteando un serio peligro para la salud de los seres vivos (Gracia-Lor, Sancho et al. 2010). Además, si no se eliminan adecuadamente en las plantas de tratamiento de aguas residuales, pueden acumularse en el medio ambiente.

Tabla 2.A. *Clasificación de los contaminantes emergentes*

Clase de Compuestos	Ejemplos
Productos Farmacéuticos Antibióticos usados en veterinaria y medicina	Trimetropim, eritromicina, lincomicina sulfametoxazol
Medicamentos analgésicos y anti-inflamatorios	Codeína, ibuprofeno, acetaminofen, ácido acetilsalicílico, diclofenaco, fenoprofen
Medicamentos psiquiátricos	Diazepam
Reguladores de lípidos	Bezafibrato, ácido clofíbrico, ácido fenofíbrico
β-bloqueantes	Metopropol, propanolol, timolol
Medios de contraste de rayos X	Iopromide, iopamidol, diatrizoato
Esteroides y hormonas (Anticonceptivos)	Estradio, estrona, estriol, dietilestilbestrol
Productos de cuidado personal Perfumes	Almizcles Nitro, policíclicos y macrocíclicos
Agentes de protección Soalr	Benzofenona, metilbenzilidene alcanfor
Repelentes de insectos	N,N-dietiltoluamida
Antisépticos	Triclosán, clorofeno
Detergentes tensoactivos y sus metabolitos	Alquilfenoles etoxilados, alquilfenoles (Nonifenol y octilfenol), alquilfenol carboxilados
Retardantes de llama	Difenil éteres polibromados (PBDEs), tetrabromo bisfenol A, Tris (2-cloroetil) fosfato
Aditivos y agentes industriales	Agentes quelantes (EDTA), sulfonatos aromáticos
Aditivos de la gasolina	Dialquil éteres, metil-t-butil éter (MTBE)
Subproductos de desinfección	Yodo-THMs, bromoácidos, bromoacetonitrilos, bromoaldehídos, cianoformaldehído, bromato, NDMA

Fuente: Barceló (Barceló 2003)

Las principales vías de entrada de estos contaminantes en el medio acuático son las aguas residuales, entre las que se incluyen las urbanas, las industriales y las agropecuarias (Barceló and Petrovic 2007). Algunos de los contaminantes emergentes pueden producir alteraciones endocrinas (ciertos pesticidas, productos químicos industriales, productos farmacéuticos y fitoquímicos, etc) afectando al sistema

hormonal e interfieriendo en los procesos reproductivos y de desarrollo de los organismos (Richardson 2009). Estos productos químicos se distribuyen extensamente en el ambiente, y pueden presentar efectos miméticos o antagónicos en el sistema endocrino alterando las funciones biológicas de las hormonas naturales, y, en consecuencia, la reproducción de los organismos afectados (Frye, Bo et al. 2012). Además, la continúa introducción de fármacos, y especialmente de los compuestos con acción antibiótica, en el medio ambiente puede provocar resistencia bacteriana.
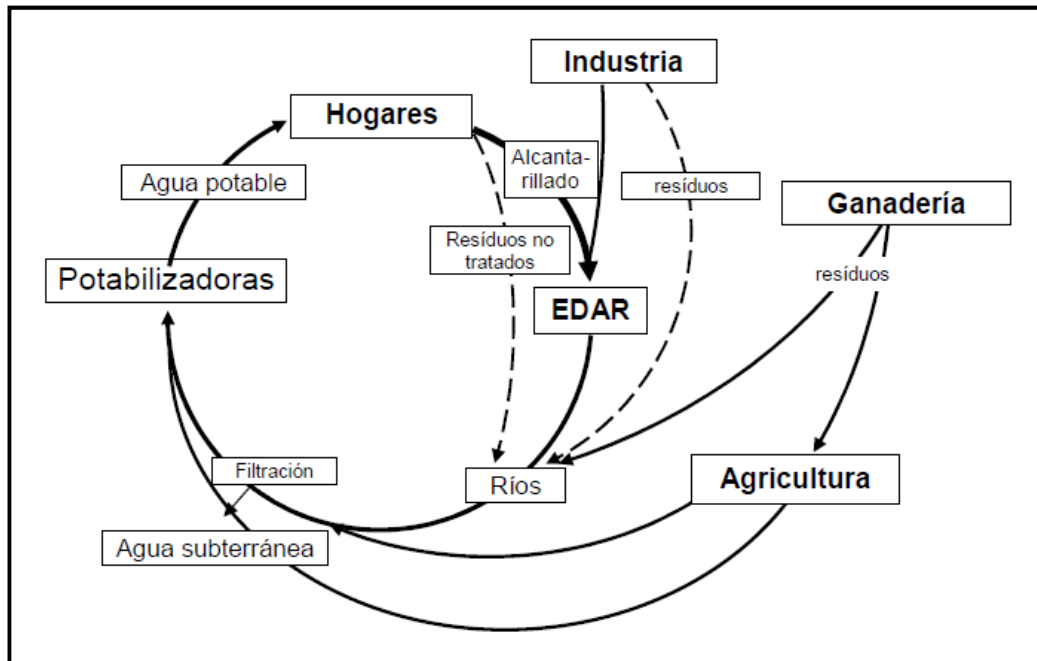
Figura 2.A. Ciclo de compartimentos ambientales del agua

Los contaminantes emergentes suelen ser compuestos con elevada solubilidad en agua, poco volátiles y de polaridad media/alta, lo que, en general, facilita su determinación por LC. Sin embargo, no se puede descartar el uso de GC, pues se trata de una técnica complementaria a LC. Debido a su heterogeneidad, sobre todo en sus propiedades químicas, resulta complicada la determinación analítica de estos compuestos mediante metodologías multiresiduales. Como ya se ha comentado en el capítulo 1, estas determinaciones suelen llevarse a cabo mediante acoplamiento LC

y/oGC a espectrómetros de masas en tándem, como QqQ o QLIT, las cuales permiten el análisis simultáneo de un número limitado de compuestos (raramente superior a 200). Además, las metodologías multiresiduales no son siempre fáciles de actualizar y ampliar (es decir, añadir nuevos compuestos a métodos ya creados), algo verdaderamente importante pues las listas de contaminantes emergentes están en continua evolución, no solo en cuanto a los compuestos inalterados sino en lo relativo a sus productos de transformación.

El desarrollo y aplicación de **métodos "universales"de screening** es, por tanto, de gran importancia en la investigación de contaminantes orgánicos en el medio ambiente, dado el elevado número de potenciales contaminantes que pueden estar presentes. Los métodos de screening deben ser rápidos y permitir, de forma fiable, una respuesta cualitativa sobre la presencia del mayor numero posible de compuestos de interés. En general, se obtiene una respuesta binaria, contestando afirmativa o negativamente a la existencia de los contaminantes incluidos en el screening, sin tener en cuenta su concentración (Portolés, Pitarch et al. 2011), aunque normalmente puede llegar a realizarse una estimación cuantitativa.

Idealmente, los métodos de screening deben cumplir diversas características. Además de ser rápidos, con poca manipulación de muestra, deben tener sensibilidad adecuada para poder controlar el cumplimiento de los límites máximos permitidos en las diferentes matrices o los niveles típicamente encontrados en las muestras.

La mayoría de contaminantes emergentes son todavía objeto de estudio. Su toxicidad y efectos están en general poco estudiados y sus productos de transformación (TPs) son muchas veces desconocidos disponiéndose de escasa información sobre su presencia en el medio ambiente. Los TPs abarcan metabolitos y productos de degradación/transformación ambiental y los generados en plantas de tratamiento de aguas residuales.Pueden tener propiedades químicas, físicas o toxicológicas muy diferentes al analito primario. Estos TPs suelen ser compuestos más polares que los productos de partida, motivo por el cual la técnica analítica usada para su determinación es LC (en detrimento de las metodologías GC), y pueden llegar a presentar una toxicidad semejante al compuesto activo (Ibáñez, Sancho et al. 2006). La degradación o transformación de los contaminantes depende de la naturaleza del

compuesto y de las condiciones ambientales. Cabe destacar que los distintos procesos de degradación, como por ejemplo la cloración o la oxidación, así como el efecto de la radiación UV, a los que puede estar sometido un determinado contaminante –y sus metabolitos- generan un amplio abanico de TPs que deben considerarse como potenciales contaminantes (Hernández, Ibáñez et al. 2008). La identificación de los productos formados es una tarea compleja que requiere del uso de tecnologías analíticas sofisticadas, fundamentalmente la cromatografía líquida acoplada a espectrometría de masas de alta resolución (Hernández, Sancho et al. 2012).

Los métodos de screening deben ser fácilmente actualizables, permitiendo ampliar fácilmente la búsqueda de nuevos contaminantes y sus TPs. Además, es conveniente que los datos obtenidos puedan ser re-evaluados de forma retrospectiva, lo cual permitiría ampliar la investigación en muestras previamente analizadas sobre la base de nueva información acerca de contaminantes y TPs que pueda aparecer en el futuro.

Para el screening pueden usarse diversas técnicas, pero las preferidas para métodos de amplio rango, por su versatilidad y sensibilidad, son las que incluyen la separación cromatográfica (GC o LC) acoplada a analizadores de masas. Los instrumentos de baja resolución (LRMS) se pueden usar con fines de screening mediante la adquisición de una transición selectiva (y sensible) para cada compuesto, en contraposición con las dos transiciones adquiridas para la identificación de un compuesto según el criterio basado en puntos de identificación (IPs) de la UE (2002/657/EC 2002). Esto permite aumentar el número de compuestos incluidos en el método, dado el menor número de transiciones medidas, aunque se requeriría un segundo análisis para la confirmación de la identidad adquiriendo transiciones adicionales para una identificación fiable del compuesto. Sin embargo, las técnicas basadas en LRMS aunque son muy sensibles, dificultan la inclusión de nuevos compuestos (necesario adquirir patrón de referencia, optimizar las condiciones de MS y MS/MS, añadir las nuevas transiciones seleccionadas) además de impedir, por su forma de trabajo, el análisis retrospectivo al no adquirir el espectro completo de masas.

La alternativa a los métodos de screening basados en instrumentos LRMS es el uso de instrumentos de alta resolución (HRMS) con los que se adquiere el espectro completo, con sensibilidad aceptable. Entre este tipo de instrumentos, se encuentran los equipos (Q)TOF y Orbitrap, los cuales permiten monitorizar, en principio, un número "ilimitado" de compuestos así como la reevaluación de los contaminantes presentes en las muestras mediante análisis retrospectivo, sin necesidad de reinyección de las mismas (Hernández, Ibáñez et al. 2011). Este tipo de analizadores ofrece además medidas de masa exacta, muy útiles en este tipo de investigaciones para la elucidación estructural y/o confirmación de la identidad de los compuestos.

La **confirmación** es objeto de debate entre la comunidad científica. Estrictamente, la confirmación requiere un reanálisis de la muestra mediante una metodología independiente a la utilizada, con el fin de asegurar que no existe ninguna contaminación que pueda llevar a un falso positivo (Lehotay, Sapozhnikova et al. 2015). La confirmación de la identidad de un compuesto (o simplemente identificación) implica asegurar que cierta respuesta obtenida es inequívocamente correspondiente a un cierto analito. Si dicha identificación no se cumple, podremos hablar simplemente de detección tentativa de un compuesto, que requiere de un segundo análisis para poder identificarlo correctamente

En los métodos de screening tienen por objeto la detección/identificación de un amplio número de compuestos para su posterior cuantificación y/o confirmación. Por ello, resulta indispensable en estos métodos definir los límites de detección del screening (SDL) y los límites de identificación (LOI). Los criterios para establecer dichos límites resultan aún confusos para las metodologías cualitativas. En el caso de técnicas MS/MS de baja resolución, los criterios son bastante uniformes, y están basados en el uso de los IPs (2002/657/EC 2002). Según este sistema, para la confirmación de la identidad en equipos LR MSMS sería necesaria la adquisición de dos transiciones al tiempo de retención del compuesto, cumpliendo las tolerancias establecidas para relaciones de intensidad Q/q (Tabla 2.B.). Para el caso de instrumentos HRMS, se requiere la adquisición de dos iones, cumpliendo también la relación Q/q. Sin embargo, las guías no establecen criterios de detección aplicables para el screening. En cuanto a la detección, en screening con analizadores LR bastaría con una

transición, mientras que con instrumentos HRMS, un único ion sería suficiente para este fin.

Tabla 2.B. *IPs conseguidos según la información obtenida en MS (izquierda) y desviación máxima tolerada de la relación Q/q en función de la abundancia relativa de ambos iones.*

MS ion	IPs
LR ion	1
LR precursor ion	1
LR product ion	1.5
HR ion	2
HR precursor ion	2
HR product ion	2.5

Abundance second ion (%)	Deviation (%)
> 50%	20
20 – 50%	25
10 – 20%	30
< 10%	50

El principal problema deriva de la falta de criterios en función de la masa exacta. Las principales guías establecen diferencias en cuanto a la resolución de los analizadores empleados, la cual suele estar estrechamente ligada a la exactitud de masa. Sin embargo, es esta última la que debería ser considerada pues es la que establece la selectividad instrumental. En nuestro grupo se propuso la obtención de estos IPs según los errores de masa obtenidos divididos en tres sectores: errores superiores a 10 mDa (LRMS), errores entre 2 y 10 mDa y errores inferiores a 2 mDa (Hernández, Ibáñez et al. 2004). Esta separación se propuso de acuerdo con las prestaciones de los equipos QTOF de hace una década, en los que los errores de masa eran más altos que los actuales. Los equipos modernos de (Q)TOF y Orbitrap ofrecen medidas de masa con errores inferiores a 1 mDa, lo que obligaría a revisar los rangos de errores de la clasificación anterior.

Por otro lado, aunque probablemente la medida óptima para discernir la capacidad confirmativa de un analizador MS sea la exactitud de masa, la resolución puede resultar crucial en el desarrollo de una metodología screening. Su importancia aumenta con la complejidad del matriz objeto de estudio, debido a la mayor

probabilidad de coelución de compuestos isobáricos. Una mayor resolución permite disminuir este efecto debido a la mayor capacidad de distinguir entre dos masas.

La complejidad de la matriz analizada no solo afecta en la aparición de falsos positivos resultantes de compuestos procedentes de dicha matriz. El **efecto matriz** (EM), asociado a los instrumentos MS y muy especialmente a los sistemas LC-MS, es sin duda el mayor problema de este acoplamiento instrumental (Marín, Gracia-Lor et al. 2009). El EM puede definirse como el aumento o disminución inesperada de la respuesta de los analitos, debido a la coelución de otros componentes presentes en la matriz y que fueron co-extraídos de la muestra (Gracia-Lor, Martínez et al. 2012). Se da principalmente durante el proceso de ionización en la interfase (APCI o ESI son más susceptibles a este afecto (Souverain, Rudaz et al. 2004), y suele ser el resultado de la competición entre los analitos o constituyentes de la matriz por la formación de los iones así como para acceder a las superficie de las gotas durante el proceso de solvatación y posterior paso al estado gaseoso (Benijts, Dams et al. 2004). Adicionalmente, algunos constituyentes pueden alterar las propiedades del eluente tales como punto de ebullición, tensión superficial y viscosidad, las cuales afectan el proceso de ionización (Bruins 1998).

Este efecto varía de una matriz a otra y depende de las características físico-químicas del analito y, sobre todo, de los interferentes que eluyen al mismo tiempo de retención que él, del ambiente en el cual se da el proceso de ionización y evaporación del ión. Dicha competencia puede generar un aumento en la señal llamado exaltación, o una disminución llamada supresión y puede afectar significativamente la precisión, sensibilidad y exactitud de un procedimiento analítico (Benijts, Dams et al. 2004).

Existen diferentes métodos para reducir o compensar el efecto matriz que permiten la correcta cuantificación de los analitos, como el método del patrón interno o el calibrado en matriz. Este efecto sobre la cuantificación no afecta a los métodos screening que únicamente persiguen la detección e identificación del compuesto. La exaltación de la respuesta puede ser, por tanto, algo beneficioso en este tipo de aproximaciones. Por otro lado, la supresión de la señal puede originar problemas en la detección de los contaminantes llevando a posibles falsos negativos por falta de

sensibilidad. La dilución de la muestra o la realización de un clean-up eficiente son las principales alternativas de reducción del EM en el análisis cualitativo, aunque ambas aproximaciones conllevan inconvenientes. La dilución está asociada a una disminución de la sensibilidad mientras que un clean-up más efectivo resulta complicado en métodos "universales", pues el comportamiento de los analitos puede ser totalmente diferente, llevando a posibles pérdidas a lo largo del proceso. Una alternativa es mejorar la eficiencia de la separación cromatográfica mediante la disminución del diámetro de partícula de la fase estacionaria, el aumento de la longitud de la columna o la mejora del gradiente de elución.

Como ya se ha comentado en el capítulo de introducción general, los métodos "universales" requieren de tratamientos de muestra sencillos y generales, con el fin de reducir los tiempos de análisis, así como minimizar las pérdidas de los analitos ya sea por la complejidad del procedimiento o por la ineficacia de éste frente a cierto tipo de analitos. Las etapas de preconcentración y eliminación de interferencias suelen ser inaplicables en el screening, a menos que exista un requerimiento de sensibilidad notable, como suele ocurrir en análisis medioambiental, en donde extracciones "universales" como SPE, LLE, etc. son comúnmente realizadas. Estas extracciones, que suelen conllevar una etapa de preconcentración, resultan de suma importancia para poder alcanzar los bajos niveles de concentración típicamente encontrados en estas muestras.

Otras matrices, como los productos alimentarios, requieren de extracción, generalmente con disolventes orgánicos, y posterior dilución para eliminar o reducir el efecto matriz y adecuar las condiciones del extracto a la cromatografía posteriormente aplicada.

Totalmente distinto es el caso de las muestras biológicas. Las muestras de orina o plasma conllevan tratamientos de muestra muy simples dado el tipo de matriz y la relativamente elevada concentración de los contaminantes o fármacos/drogas en ellas. Una simple centrifugación (en algunos casos tras dilución) suele ser suficiente. Otras muestras biológicas más complejas por la elevada cantidad de grasas que contienen, como puede ser las de pulmón, hígado, etc., son generalmente muy invasivas y poco utilizadas para la búsqueda de tóxicos.

---------------------- o ------------------------

En este capítulo se exploran las capacidades y limitaciones de los instrumentos LC-(Q)TOF MS, o más concretamente el acoplamiento UHPLC-(Q)TOF MS, para el screening de contaminantes, residuos y drogas en diferentes matrices de interés ambiental, alimentario y toxicológico. En primer lugar, ha sido necesaria la creación de bases de datos y librerías de espectros tanto teóricas como experimentales. Las primeras se basan en listados de compuestos previamente reportados en la literatura y analizados por LC-MS bajo condiciones relativamente generales, es decir sin necesidad de fases móviles, aditivos o columnas cromatográficas específicas ni derivatización u otros procesos similares. Éstas bases de datos teóricas incluyen la formula molecular de cada compuesto y la masa exacta de la molécula (des)protonada, así como de algunos aductos en base a la posibilidad de que estos pudieran formarse de forma importante, bien sea por un conocimiento previo o por que hubiera sido previamente reportado. En cuanto a la librería experimental de espectros, requiere del análisis de patrones y muestras bajo condiciones predefinidas y constantes, de tal manera que tanto la información espectral como los tiempos de retención presenten variaciones mínimas. Este el es objetivo del **artículo científico 1** en el que se estudian y optimizan los diferentes parámetros que pueden afectar a la calidad y robustez de la librería de espectros experimental, centrándose sobre todo en aquellos que afectan a la sensibilidad, exactitud de masa y robustez de la señal. Se estudian además diferentes alternativas para la fragmentación de los iones, con la finalidad de maximizar la información obtenida en una única inyección y realizar una confirmación de identidad de forma automática.

Una vez definidos las variables de la metodología UHPLC-(Q)TOF MS para el screening, se estudian diferentes aproximaciones para el tratamiento de los datos obtenidos y la búsqueda de contaminantes. Existen dos posibilidades que cumplen con los requisitos del screening: post-target y non-target. La segunda aproximación puede hacer uso de librerías de espectro teóricas y empíricas para la detección de los componentes/compuestos encontrados. Así pues, en el **artículo científico 2** se comparan las ventajas y limitaciones de las diferentes estrategias en base a los

resultados obtenidos en muestras reales de aguas residuales depuradas (efluente urbano, EWW), vegetales (naranja, plátano y maíz) y orina de voluntarios sometidos a un tratamiento de desintoxicación. En este artículo se discuten los diferentes aspectos que caracterizan a las aproximaciones post y non-target en base a resultados reales. Además, se estudia la aplicabilidad de la técnica UHPLC-(Q)TOF MS para llevar a cabo un screening "universal" en muestras de origen muy diverso.

Finalmente, en el **artículo científico 3**, se efectúa una validación de tipo cualitativo en aguas, considerando tres tipos diferentes: subterránea, superficial y EWW. La validación se ha llevado a cabo para aquellos contaminantes de los cuales se dispone de patrón de referencia y se basa en una respuesta binaria SI/NO se detecta el contaminante estudiado en las diferentes matrices a los dos niveles de concentración evaluados (0.1 y 1 µg/L), obteniendo así el SDL para cada compuesto. De igual manera, se define el LOI en base a la respuesta binaria SI/NO se identifica el analito. En este trabajo se definen y debaten también los conceptos de confirmación e identificación en el screening mediante instrumentación HRMS en adquisición full scan y, sobretodo, cuando se hace uso de técnicas como el MS^E o similares.

2.2. Artículo científico 1

"Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry"

R. Díaz, M. Ibáñez, J. V. Sancho and F. Hernández.

Rapid Communications in Mass Spectrometry, 25 (2011) 355-369.

Research Article

Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/ hybrid quadrupole time-of-flight mass spectrometry

Ramón Díaz, María Ibáñez, Juan V. Sancho and Félix Hernández*

Research Institute for Pesticides and Water, University Jaume I, E-12071 Castellón, Spain

Hybrid quadrupole time-of-flight mass spectrometry (QTOF MS) has gained wide acceptance in many fields of chemistry, for example, proteomics, metabolomics and small molecule analysis. This has been due to the numerous technological advances made to this mass analyser in recent years. In the environmental field, the instrument has proven to be one of the most powerful approaches for the screening of organic pollutants in different matrices due to its high sensitivity in full acquisition mode and mass accuracy measurements. In the work presented here, the optimum experimental conditions for the creation of an empirical TOF MS spectra library have been evaluated. For this model we have used a QTOF Premier mass spectrometer and investigated its functionalities to obtain the best MS data, mainly in terms of mass accuracy, dynamic range and sensitivity. Different parameters that can affect mass accuracy, such as lock mass, ion abundance, spectral resolution, instrument calibration or matrix effect, have also been carefully evaluated using test compounds (mainly pesticides and antibiotics). The role of ultra-high-pressure liquid chromatography (UHPLC), especially when dealing with complex matrices, has also been tested. In addition to the mass accuracy measurements, this analyser allows the simultaneous acquisition of low and high collision energy spectra. This acquisition mode greatly enhances the reliable identification of detected compounds due to the useful (de)protonated molecule and fragment ion accurate mass information obtained when working in this mode. An in-house empirical spectral library was built for approximately 230 organic pollutants making use of QTOF MS in MS^E mode. All the information reported in this paper is made available to the readers to facilitate screening and identification of relevant organic pollutants by QTOF MS. Copyright © 2010 John Wiley & Sons, Ltd.

Orthogonal-acceleration time-of-flight (oaTOF) and hybrid quadrupole oaTOF (QTOF) mass spectrometers have been valuable tools in many fields of chemistry including impurity profiling of pharmaceutical drug substances,[1] metabolomics[2,3] and food safety.[4–11] Recently, they have also gained wide acceptance in the field of environmental analysis for the rapid screening of organic pollutants due to their high sensitivity in full acquisition mode and their ability to provide accurate mass measurements.[12–14] QTOF MS has also been applied for the elucidation of unknowns because of its capacity to establish reliable elemental compositions.[15–17] However, although QTOF mass spectrometers have been commercially available for over ten years, they have had drawbacks which have reduced their analytical application. These include the narrow ion abundance range over which accurate mass measurements could be made with a high degree of certainty, and the low dynamic range.[18] However, thanks to continuous development of the technology, the latest TOF analysers show considerable advancement over earlier models.

Most significant enhancements are related to mass measurement accuracy, resolution, dynamic range, sensitivity and speed. One of the major problems with TOF MS analysis is the saturation of the time-to-digital converter (TDC) detectors due to their inherent dead time, which affects not only the linearity, but also the mass accuracy.[19] To circumvent this problem, an analogue-to-digital converter (ADC) detector could be used, as it digitises the ion current from the output of the micro-channel plate (MCP). The use of an ADC device avoids the saturation effects and allows the ion detector to render an accurate count of ions even for abundant species, although there is an inherent background noise associated with analogue detection.[19] Another option, the one used in our instrument, is a travelling wave ion guide (TWIG), which produces an attenuated ion beam, reducing the number of ions that reach the detector, and thus avoiding detector saturation. This results in a substantial increase in dynamic range for accurate mass measurement,[20] known as programmable Dynamic Range Enhancement (pDRE). Wide dynamic ranges are necessary for the analysis of unknown samples or for samples that vary widely in analyte concentrations. Otherwise, many samples would require a reanalysis after dilution, generating additional work and increasing the time needed to obtain reliable data.[21] Furthermore, TWIG also permits the instrument duty cycle to be significantly enhanced (Enhanced Duty Cycle, EDC),

* *Correspondence to:* F. Hernández, Research Institute for Pesticides and Water, University Jaume I, E-12071 Castellón, Spain.
E-mail: felix.hernandez@qfa.uji.es

355

Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry

R. Díaz, M. Ibáñez, J.V. Sancho and F. Hernández

Research Institute for Pesticides and Water, Universitat Jaume I, E-12071 Castelló, Spain

ABSTRACT

Hybrid quadrupole time-of-flight mass spectrometry (QTOF MS) has gained wide acceptance in many fields of chemistry, for example, proteomics, metabolomics and small molecule analysis. This has been due to the numerous technological advances made to this mass analyser in recent years. In the environmental field, the instrument has proven to be one of the most powerful approaches for the screening of organic pollutants in different matrices due to its high sensitivity in full acquisition mode and mass accuracy measurements. In the work presented here, the optimum experimental conditions for the creation of an empirical TOF MS spectra library have been evaluated. For this model we have used a QTOF Premier mass spectrometer and investigated its functionalities to obtain the best MS data, mainly in terms of mass accuracy, dynamic range and sensitivity. Different parameters that can affect mass accuracy, such as lock mass, ion abundance, spectral resolution, instrument calibration or matrix effect, have also been carefully evaluated using test compounds (mainly pesticides and antibiotics). The role of ultra-high-pressure liquid chromatography (UHPLC), especially when dealing with complex matrices, has also been tested. In addition to the mass accuracy measurements, this analyser allows the simultaneous acquisition of low and high collision energy spectra. This acquisition mode greatly enhances the reliable identification of detected compounds due to the useful (de)protonated molecule and fragment ion accurate mass information obtained when working in this mode. An in-house empirical spectral library was built for approximately 230 organic pollutants making use of QTOF MS in MSE mode. All the information reported in this paper is made available to the readers to facilitate screening and identification of relevant organic pollutants by QTOF MS.

INTRODUCTION

Orthogonal-acceleration time-of-flight (oaTOF) and hybrid quadrupole oaTOF (QTOF) mass spectrometers have been valuable tools in many fields of chemistry including impurity profiling of pharmaceutical drug substances,[1] metabolomics[2, 3] and food safety.[4–11] Recently, they have also gained wide acceptance in the field of environmental analysis for the rapid screening of organic pollutants due to their high sensitivity in full acquisition mode and their ability to provide accurate mass measurements.[12–14] QTOF MS has also been applied for the elucidation of unknowns because of its capacity to establish reliable elemental compositions.[15–17] However, although QTOF mass spectrometers have been commercially available for over ten years, they have had drawbacks which have reduced their analytical application. These include the narrow ion abundance range over which accurate mass measurements could be made with a high degree of certainty, and the low dynamic range.[18] However, thanks to continuous development of the technology, the latest TOF analysers show considerable advancement over earlier models.

Most significant enhancements are related to mass measurement accuracy, resolution, dynamic range, sensitivity and speed. One of the major problems with TOF MS analysis is the saturation of the time-to-digital converter (TDC) detectors due to their inherent dead time, which affects not only the linearity, but also the mass accuracy.[19] To circumvent this problem, an analogue-to-digital converter (ADC) detector could be used, as it digitises the ion current from the output of the micro-channel plate (MCP). The use of an ADC device avoids the saturation effects and allows the ion detector to render an accurate count of ions even for abundant species, although there is an inherent background noise associated with analogue detection.[19] Another option, the one used in our instrument, is a travelling wave ion guide (TWIG), which produces an attenuated ion beam, reducing the number of ions that reach the detector, and thus avoiding detector saturation. This results in a substantial increase in dynamic range for accurate mass measurement,[20] known as programmable Dynamic Range Enhancement (pDRE). Wide dynamic ranges are necessary for the analysis of unknown samples or for samples that vary widely in analyte concentrations. Otherwise, many samples would require a reanalysis after dilution, generating additional work and increasing the time needed to obtain reliable data.[21] Furthermore, TWIG also permits the instrument duty cycle to be significantly

enhanced (Enhanced Duty Cycle, EDC), hence achieving higher sensitivity for a limited and preselected range of the mass spectrum.[22]

Another feature implemented in TOF MS analysers is the dedicated reference sprayer in the mass spectrometer source. Early instruments introduced the reference compound(s) together with the mobile phase in a post-column T. This approach presented a significant drawback as it led to matrix ionization suppression, decreasing the reference compound sensitivity and increasing mass errors. There have been different proposals to improve the robustness of the exact mass measurements. One, using an additional electrospray source that orthogonally generates lock mass ions separately, allows good mass accuracies to be obtained independently of the sample conditions.[23] Frequency and scan time for lock mass acquisition can be chosen by the user. When using ultra-high-pressure liquid chromatography (UHPLC) in combination with TOF MS, the narrow chromatographic peaks achieved make the overlapping between lock mass acquisition and analyte peak lower. However, when this occurs, the impact on peak shape may be significant.

In this paper, we are particularly interested in assessing the parameters that contribute to the accuracy and precision of mass measurements using the QTOF with UHPLC as the separation technique. Resolution, dynamic range and sensitivity are also important parameters to be considered. The main parameters studied are those that can affect the mass accuracy, such as the acquisition interval and the scan time for lock mass or pDRE. The sensitivity improvement when increasing the duty cycle was also investigated as well as its influence on mass accuracy. We have also evaluated the influence of the UHPLC technique on matrix effects when dealing with the analysis of complex matrices, as this is a key parameter in accurate mass measurements.[24] The results have been compared with those obtained in W acquisition mode (the flight path length is doubled resulting in almost twice the mass resolving power). This is particularly relevant when isobaric compounds are coeluting (more likely in highly complex matrices). To date only a few papers have been published on testing the capabilities of MS instruments,[18, 21, 25, 26] and these have mostly been based on the analysis of a few model compounds. In order to obtain more complete knowledge of QTOF MS capabilities in environmental analysis, it would be desirable to increase the number of compounds evaluated as well as the range of their physicochemical characteristics.

Full spectrum MS data is obtained when using TOF MS. Therefore, when performing screening analyses, the contaminants present in samples must be extracted from the Total Ion

Chromatogram (TIC). There are two main ways that can be considered when screening by LC/high resolution MS in environmental analysis:[14, 17, 27] target screening searching for a list of suspected compounds of interest (with or without reference standards), and non-target screening that would ideally require a mass spectral library to rapidly and automatically identify those components found by the algorithm applied. Otherwise an extensive, tedious investigation would be required in non-target screening with the aim of assigning reliable empirical formulae to each peak extracted from the TIC, and searching afterwards for potential chemical structures in compound databases. Unlike GC/EI-MS, commercial mass spectral libraries are not yet available in LC/API-MS, mainly due to the non-availability of a standardized interface. Currently, the differences in the ionization process between the existing interfaces, together with the variability in the results depending on the mobile phase composition or in-source voltages applied, make it difficult to use standardized libraries.[28] Therefore, building home-made mass spectral libraries would facilitate searching, with particular focus on relevant compounds. Obviously, the higher the number of compounds included in the library, the greater the possibility of detecting and identifying more contaminants in samples.

After the optimization of the above-mentioned instrumental parameters, we studied the optimum conditions for building an empirical TOF MS spectral library, trying to establish the most appropriate fragmentation approach to improve compound identification. When using soft ionisation sources (e.g. electrospray) relatively poor spectral information is obtained compared with GC/EI-MS, because only the (de)protonated molecule and/or some adduct ions are normally present. Promoting analyte fragmentation is advantageous in greatly increasing the number of characteristic ions that can then be matched against the empirical spectral library. In this study, we have optimised the capabilities of the QTOF instrument, with the objective of enhancing its efficiency and reliability when performing screening analyses of real-world samples.

EXPERIMENTAL

Reagents and chemicals

Around 40 analytes were studied for performance optimisation and a total of 230 compounds were included in the empirical mass spectral library (see Table 1). Reference compounds were purchased from Acros Organics (Geel, Belgium), Bayer Hispania (Barcelona, Spain), Fort Dodge Veterinaria (Gerona, Spain), Vetoquinol Industrial (Madrid, Spain), Aventis

Pharma (Madrid, Spain), Sigma Aldrich (St Louis, MO, USA), Cerilliant (Round Rock, TX, USA), Dr. Ehrenstorfer (Augsburg, Germany), Riedel-de Haën (Seelze, Germany), the National Measurement Institute (Pymble, Australia) and Fluka (Buchs, Switzerland). All reference materials had purities higher than 93% (w/w).

Table 1. List of compounds included in the empirical spectral library and accurate mass measurements for up to five most abundant ions (higher than 10%) in the HE spectra (normalised ion abundance in brackets). RT was obtained using an Acquity UPLC BEH C_{18} 1.7 µm particle size analytical column, 150 x 2.1 mm (mobile phases and gradient conditions are explained in the Experimental section). The main product ions in the LE function are marked with *

Compound	RT (min)	*Exact mass[c] [M + H]+*	*Exact mass[c] [M-H]−*	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
1,4,6-Androstatrien-17-ol-3-one	10.66	285.1854		147.0797 (100)	121.0644 (70)	171.0806 (55)	285.1849 (30)	119.0856 (26)
1,4-Androstadienedione	10.24	285.1854		121.0643 (100)	–	–	–	–
11-Nor-9-carboxy-Δ9-THC	14.46	345.2066		299.2095 (16)	–	–	–	–
16-Hydroxy-stanozolol	10.93	345.2542		345.2547 (100)	–	–	–	–
19-Hydroxy-androstenedione	8.58	303.1960		303.1956 (100)	157.1017 (50)	267.1761 (42)	255.1757 (42)	145.1018 (42)
1-Testosterone	12.01	289.2167		187.1485 (100)	289.2153 (41)	105.0693 (35)	131.0851 (35)	91.0534 (33)
2-Hydroxy-atrazine	5.07	198.1355		86.0349 (100)	156.0877 (100)	114.0663 (77)	97.0401 (43)	69.0094 (38)
2-Hydroxy-simazine	3.75	184.1198		114.0659 (100)	69.0094 (44)	97.0389 (34)	184.1207 (21)	–
2-Hydroxy-terbutylazine	6.37	212.1511		156.0840 (100)	86.0348 (83)	114.066 (54)	97.0399 (41)	69.0096 (39)
3,4-Dichloraniline	9.31	161.9877		127.0186 (100)	129.0160 (32)	108.9843 (17)	–	–
3-Hydroxy-carbofuran	6.00	238.1079		107.0494 (100)	135.0806 (60)	163.0761 (38)	145.0650 (14)	–
3'-Hydroxy-stanozolol	11.15	345.2542		345.2552	–	–	–	–

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
				(100)				
4,6-Androstadien-17-ol-3-one	11.17	287.2011		133.1000 (100)	105.0693 (70)	287.1997 (59)	151.1110 (48)	97.0647 (33)
4,9-Androstadien-17-Me-17-ol-3-one	11.65	301.2168		301.2155 (100)	283.2029 (82)	147.0801 (81)	159.1165 (63)	253.1579 (53)
4,9-Androstadien-17-ol-3-one	11.15	287.2011		147.0798 (100)	287.1998 (70)	105.0695 (50)	145.1008 (45)	269.1895 (44)
4-Aminoantipyrin	5.34	204.1137		83.0614 (100)	94.0649 (91)	161.0714 (67)	159.0912 (55)	104.0496 (47)
4-Hydroxy-stanozolol	11.17	345.2542		145.0763 (100)	309.2326 (77)	327.2434 (59)	133.0766 (34)	147.0930 (31)
4-Hydroxy-testosterone	11.45	305.2116		125.0594 (100)	113.0594 (89)	305.2091 (34)		
6b-Hydroxy-testosterone	8.08	305.2116		305.2120 (100)	269.1913 (51)	157.10022 (43)	145.1017 (38)	–
6-Hydroxy-4-Cl-dehydromethyltestosterone	9.91	351.1727		155.0262 (71)	105.0700 (69)	147.1171 (60)	91.0540 (43)	315.1526 (61)[*]
6-Hydroxy-androstenedione	8.98	303.1960		303.1965 (46)	97.0649 (41)	285.1855 (36)	267.1746 (26)	
Acetaminophen	2.70	152.0711		150.0562 (100)	134.0609 (34)	110.0611 (34)	93.0347 (19)	92.0503 (12)
Acetamiprid	5.96	223.0750		126.0113 (100)	128.0087 (35)	–	–	–
Acrinathrin	15.30	542.5746		–	–	–	–	–
Aflatoxin B1	8.19	313.0712		241.0497 (14)	285.0757 (11)	269.0450 (10)	–	–
Aflatoxin B2	7.79	315.0868		337.0706 (100)	315.0830 (18)	–	–	–
Aflatoxin G1	7.36	329.0661		351.0491 (100)	–	–	–	–
Aflatoxin G2	6.93	331.0818		353.0646 (100)	331.0791 (18)	–	–	–
Aflatoxin M1	7.02	329.0661		273.0771 (40)	259.0628 (21)	229.0507 (19)	283.0624 (11)	227.0706 (11)
Alachlor[a]	12.22	270.1261		147.1047 (100)	162.1279 (85)	132.0818 (81)	146.0971 (48)	90.0109 (42)

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Aldicarb	7.30	191.0854		213.0679 (100)*	89.0420 (61)*	116.0533 (35)*	–	–
Aldicarb sulfone	3.48	223.0753		118.9484 (100)	86.0603 (59)	109.0506 (48)	113.9871 (36)	103.9251 (28)
Aldicarb sulfoxide	3.19	207.0803		89.0425 (100)	113.9867 (16)	89.0416 (100)*	229.0628 (62)*	166.0721 (50)*
Alprazolam	10.32	309.0907		281.0707 (100)	205.074 (61)	283.0740 (42)	309.0899 (40)	274.1216 (25)
Amphetamine	3.06	136.1126		91.0554 (14)*	119.0867 (14)*	–	–	–
Ampicillin[a]	4.52	350.1174		106.0657 (100)	160.0431 (22)	114.0011 (21)	118.0652 (20)	350.1158 (19)
Atorvastatin	12.40	559.2608		440.2204 (53)	250.1026 (23)	292.1508 (15)	–	
Atrazine[a]	9.86	216.1016		174.0507 (100)	96.0556 (97)	104.0012 (82)	132.0324 (47)	79.0068 (46)
Azinphos-methyl	10.49	318.0136		132.0447 (100)	104.0497 (32)	105.0339 (31)	77.0391 (12)	–
Azoxystrobin	10.97	404.1246		329.0779 (93)	344.1003 (55)	372.0938 (38)	–	–
Benfuracarb	14.05	411.1954		186.0659 (100)	195.0495 (28)	102.0013 (14)	213.0822 (11)	–
Bensulide	12.60	398.0683		240.0114 (100)	141.0003 (85)	158.0270 (59)	77.0389 (33)	–
Bentazone	7.30		239.0491	239.0472 (100)	132.0317 (84)	197.0009 (32)	175.0866 (25)	–
Benzoylecgonine	5.06	290.1392		168.0984 (100)	105.0324 (95)	82.0649 (39)	119.0490 (22)	150.0916 (19)
Bezafibrate	11.06	362.1159		138.9951 (100)	121.0654 (73)	298.0598 (33)	140.9930 (27)	314.0259 (27)
Bifenazate	11.92	301.1552		170.0966 (100)	152.0623 (90)	167.0741 (49)	156.0824 (35)	184.0789 (33)
Boldenone metabolite	12.14	289.2168		187.1479 (100)	289.2158 (73)	145.1011 (55)	69.0339 (54)	131.0853 (52)
Bromacil	8.35	261.0239		204.9606 (100)	206.9596 (98)	–	–	–
Buprofezin[a,b]	14.25	306.1640		106.0648	86.0593	145.0424	–	–

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
				(100)	(29)	(11)		
Butocarboxim	7.18	191.0819		75.0273 (100)	–	–	–	–
Carbaryl	9.03	202.0813		127.0539 (100)	145.0649 (65)	–	–	–
Carbendazim[a,b]	5.24	192.0773		160.0481 (100)	132.056 (35)	105.0454 (15)	–	–
Carbofuran[a,b]	8.61	222.1130		123.0447 (100)	165.0923 (11)	–	–	–
Cefotaxim[a]	4.52	456.0647		125.0032 (100)	126.0124 (76)	156.0216 (59)	167.0274 (54)	140.0174 (42)
Cephalexin[a]	4.29	348.1018		91.0545 (100)	106.0654 (93)	118.0655 (75)	158.0276 (69)	174.0563 (49)
Chloramphenicol	6.46	323.0201		165.0662 (100)	119.0728 (43)	275.0003 (37)	–	–
Chlorpyriphos	14.60	349.9341		197.9283 (100)	199.9256 (98)	96.9511 (89)	134.899 (50)	201.9237 (34)
Chlorpyriphos-methyl	13.44	321.9027		124.9827 (100)	93.0111 (35)	78.9964 (21)	109.0041 (17)	–
Chlorsulfuron	8.95	358.0376		141.0770 (100)	163.0594 (96)	167.0564 (87)	127.0538 (38)	179.0265 (34)
Chlortetracyclin[a]	5.07	479.1221		462.0935 (100)	464.0944 (33)	196.9999 (15)	–	–
Ciprofloxacin[a]	4.23	332.141		231.0565 (100)	314.1253 (49)	245.1076 (28)	203.0623 (17)	288.1484 (15)
Clarithromycin[a]	10.34	748.4847		158.1175 (100)	590.3905 (17)	83.0489 (14)	–	–
Clindamycin[a]	7.82	425.1877		126.1276 (100)	–	–	–	–
Clothianidin	5.24	250.0165		131.9676 (100)	110.0717 (56)	113.0167 (47)	133.9653 (36)	169.0552 (29)
Cloxacillin[a]	9.93	436.0734		182.0249 (100)	299.0203 (95)	160.0429 (68)	178.0067 (53)	277.0381 (47)
Cocaethylene	6.14	318.1705		196.1302 (100)	82.0642 (65)	105.0331 (29)	150.0914 (25)	–
Cocaine	4.97	304.1549		182.1138 (100)	82.0656 (58)	105.0336 (54)	150.0920 (28)	119.0498 (23)

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Codeine	2.82	300.1599		300.1593 (100)	215.1089 (25)	165.0713 (23)	199.0773 (22)	225.0929 (17)
Coumaphos	12.91	363.0223		226.9937 (100)	228.9923 (39)	211.0172 (22)	288.9503 (19)	306.9609 (16)
Cyprodinil	12.53	226.1344		226.1335 (100)	93.0563 (33)	108.0789 (29)	118.0554 (19)	133.0756 (15)
Deethylatrazine	6.62	188.0703		146.0229 (100)	104.0008 (61)	79.0065 (40)	110.0466 (34)	148.0195 (31)
Deethylterbumetone	7.97	198.1355		142.0723 (100)	86.0351 (57)	100.0512 (21)	–	–
Deethylterbuthylazine	8.93	202.0890		146.0205 (100)	104.0013 (46)	148.0202 (43)	79.0067 (34)	110.0467 (29)
Deisopropylatrazine (DIA)	4.87	174.0546		104.0013 (100)	96.0559 (93)	68.0255 (85)	79.0067 (58)	132.0300 (46)
Deltamethrin	15.36	503.9810		–	–	–	–	–
Deoxynivalenol	3.72	297.1338		187.0773 (100)	225.0237 (63)	175.0773 (61)	287.0612 (60)	205.0879 (46)
DHT	12.40	291.2324		255.2106 (100)	159.1168 (82)	291.2320 (82)	145.1012 (70)	119.0854 (54)
Diazinon	13.05	305.1089		169.0795 (100)	153.1019 (54)	96.9509 (32)	100.0219 (20)	84.0444 (10)
Dichlovos (DDVP)	8.45	220.9537		–	–	–	–	–
Diclofenac	12.59	296.0245		214.0421 (100)	216.0418 (35)	–	–	–
Dicloxacillin[a]	10.54	470.0344		182.0251 (100)	332.981 (79)	334.9802 (54)	–	–
Diethofencarb	11.02	268.1549		124.0382 (100)	152.0690 (25)	108.0447 (13)	–	–
Diflubenzuron	12.32	311.0399		141.0146 (100)	195.9887 (40)	158.0416 (22)	140.0310 (20)	
Diflufenican	5.46	395.0819		266.0443 (100)	246.0377 (24)	238.0505 (14)	–	–
Dimethomorph (2 stereoisomers)	11.27	388.1316		165.0548 (100)	301.0622 (95)	303.0616 (36)	138.9960 (18)	–
	11.66			165.0547 (100)	301.0624 (93)	303.0629 (36)	138.9957 (20)	–

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Dimethoate[a]	5.76	230.0075		124.9825 (100)	78.9957 (18)	93.0101 (17)	–	–
Diphenylamine	11.79	170.0970		93.0576 (100)	170.0969 (21)	–	–	–
Diuron[a]	9.96	233.0248		72.045 (100)	161.0709 (14)	–	–	–
Enalapril	8.77	377.2076		234.1479 (100)	117.0701 (84)	160.1126 (74)	130.0869 (57)	91.0547 (28)
Epoxyconazole	12.19	330.0809		121.0450 (100)	123.0235 (30)	101.0384 (10)	–	–
Erythromycin[a]	9.41	734.469		158.1182 (100)	576.3747 (24)	–	–	–
Erythromycin - H2O	9.87	716.4584		158.1172 (100)	558.3623 (16)	116.1067 (15)	540.3533 (11)	522.3424 (10)
Ethiofencarb	9.36	226.0902		107.0487 (100)	147.0143 (15)	–	–	–
Ethiofencarb sulfone	5.11	258.0800		107.0489 (100)	79.0542 (13)	–	–	–
Ethiofencarb sulfoxide	5.35	242.0851		107.0490 (100)	77.0391 (16)	79.0546 (13)	–	–
Ethofenprox	16.15	377.2116		183.0816 (100)	107.0496 (39)		–	–
Ethofumesate	11.01	287.0953		161.0602 (100)	121.0647 (65)	133.0650 (62)	137.0595 (29)	115.0553 (23)
Ethoxazole	14.90	360.1775		141.0143 (100)	–	–	–	–
Ethoxyquin	11.29	218.1545		174.0917 (100)	160.0754 (27)	202.1213 (21)	148.0748 (21)	218.1492 (19)
Fenarimol	12.13	331.0405		268.0528 (100)	81.0445 (76)	138.9948 (73)	189.0706 (59)	259.0098 (37)
Fenhexamid	12.11	302.0714		97.1007 (100)	302.0713 (14)	304.0481 (10)	–	–
Fenoxaprop	12.66	334.0482		262.0291 (50)	288.0435 (44)	356.0192 (36)	373.9948 (36)	261.0212 (25)
Fenoxycarb	12.59	302.1392		88.0392 (100)	–	–	–	–
Fenthion	12.82	279.0278		105.0698	169.0132	109.0058	153.0360	138.0504

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
				(100)	(97)	(29)	(28)	(24)
Fluazifop-p-butyl	14.16	384.1422		282.0703 (100)	238.0482 (61)	254.0629 (38)	91.0533 (34)	255.0559 (30)
Fludioxonil	11.28	249.0475		158.0389 (100)	131.0289 (23)	185.0507 (22)	165.0457 (14)	138.0347 (11)
Flufenoxuron	14.90	489.0440		158.0419 (100)	141.0148 (36)	312.0059 (11)	–	–
Flumequine	8.60	262.0879		202.0306 (100)	244.0762 (82)	174.0358 (17)	–	–
Flutriafol	9.97	302.1105		123.0241 (100)	70.0401 (68)	109.0442 (34)	95.0290 (11)	–
Formebolone metabolite	9.17	347.2222		539.2966 (48)	147.0802 (25)	347.2234 (20)		–
Fosmet	10.59	318.0023		160.0393 (100)	133.0286 (21)	105.033 (13)	–	–
Fumonisin B1	10.13	722.3963		722.3948 (100)	704.3864 (27)	334.3112 (23)	352.3232 (23)	–
Fumonisin B2	11.66	706.4014		706.3996 (100)	336.3272 (42)	688.3908 (25)	318.3172 (24)	354.3374 (17)
Furazolidone[a]	3.84	226.0464		122.0109 (100)	139.0137 (36)	67.042 (35)	95.0368 (35)	79.0182 (20)
Furoxypir	7.18	254.9739		208.9693 (100)	210.9672 (66)	254.9760 (43)	236.9648 (42)	234.9178 (32)
Gemfibrozil	13.81	251.1647		167.0389 (46)	–	–	–	–
Gestrinone	11.33	309.1854		199.1108 (100)	309.1841 (87)	241.1586 (75)	262.1367 (47)	291.1748 (35)
GHB	2.41	105.0552		127.0370 (100)	–	–	–	–
Haloxyfop	14.21	362.0407		–	–	–	–	–
Haloxyfop-2-ethoxyethyl	14.15	434.0982		316.0354 (100)	272.0117 (48)	288.0417 (42)	318.0345 (39)	274.0098 (15)
Haloxyfop-methyl	13.68	376.0563		272.0101 (100)	316.0342 (88)	288.0355 (66)	91.0534 (48)	274.0093 (34)
Hexythiazox	14.64	353.1091		168.0583 (100)	116.0618 (79)	413.2675 (70)	115.0539 (62)	151.0315 (41)

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
HT-2 toxin	10.16	425.2175		345.132 (31)	285.1115 (25)	–	–	–
Imazalil[a,b]	9.12	297.0561		158.978 (100)	160.9749 (71)	69.0451 (28)	172.9938 (17)	81.0448 (16)
Imidacloprid[a,b]	5.17	256.0601		209.0579 (100)	175.0962 (97)	211.0587 (36)	173.0831 (33)	210.0648 (21)
Iprodione	12.49	330.0412		282.9355 (100)	284.9361 (35)	239.9370 (11)	–	–
Isoproturon[a]	9.99	207.1497		72.0454 (100)	–	–	–	–
Ketamine	4.90	238.0998		125.0153 (100)	127.0137 (35)	163.0322 (15)	165.0211 (11)	220.0898 (10)
Ketoprofen	10.57	255.1021		105.0341 (100)	209.0958 (29)		–	–
Lincomycin[a]	3.79	407.2216		126.1278 (100)	–	–	–	–
Linuron	10.95	249.0197		159.9723 (100)	132.9616 (94)	160.9805 (79)	161.9703 (72)	182.0246 (70)
Lorazepam	10.22	321.0197		275.0138 (100)	277.011 (69)	229.053 (56)	163.0068 (24)	–
Lufenuron	14.33	510.9862		–	–	–	–	–
Malathion[a]	11.42	331.0439		124.9826 (100)	99.0081 (95)	78.9949 (14)	71.0144 (11)	–
Marbofloxacin	3.43	363.1468		72.0812 (100)	205.0419 (14)	345.135 (13)	–	–
MCPA	9.88		199.0162	141.0107 (100)	143.0087 (37)	199.0163 (32)	201.0135 (10)	–
MDA	3.08	180.1024		105.0692 (100)	77.0396 (68)	79.0552 (58)	135.0440 (57)	103.0552 (39)
MDEA	3.13	208.1338		105.0683 (100)	135.0423 (61)	77.0384 (55)	79.0540 (49)	133.0636
MDMA	3.14	194.1180		105.0683 (100)	135.0423 (61)	77.0384 (55)	79.0540 (49)	133.0636 (35)
Mesotrione	6.03	340.0491		227.9957 (100)	104.013 (51)	362.0238 (37)	293.0493 (22)	249.9957 (19)
Mesterolone	12.75	305.2480		305.2480 (100)	173.1327 (66)	229.1953 (60)	95.0853 (57)	269.2278 (56)

Compound	RT (min)	*Exact mass*[c] *[M + H]*[+]	*Exact mass*[c] *[M-H]*[−]	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Methamidophos	2.09	142.0091		94.0043 (100)	142.0072 (100)*	112.0141 (37)*	124.9801 (37)*	94.0037 (22)*
Methamphetamine	3.15	150.1282		91.0537 (100)	91.0549 (100)*	150.1290 (52)*	119.0864 (18)*	–
Methidathion	10.29	302.9697		85.039 (100)5	145.0076 (49)	–	–	–
Methiocarb	11.09	226.0902		121.0648 (100)	122.073 (66)	77.0398 (59)	107.0495 (53)	91.0545 (41)
Methiocarb sulfone	6.27	258.0800		122.0715 (100)	107.049 (62)	–	–	–
Methiocarb sulfoxide	5.73	242.0851		122.0719 (100)	170.0399 (58)	153.0368 (46)	168.0604 (41)	107.049 (38)
Methomyl	4.05	163.0541		88.0215 (100)	113.9863 (65)	72.9989 (29)	141.982 (12)1	128.0123 (10)
Methyl-1-testosterone	12.54	303.2324		201.1619 (100)	145.1009 (84)	303.2312 (48)	105.0697 (37)	159.1171 (34)
Metolachlor[a]	12.34	284.1417		176.1423 (100)	252.112 (58)	134.0966 (38)	254.1116 (25)	–
Metribuzin	8.37	215.0966		187.1017 (100)	215.0947 (60)	74.0069 (30)	89.0162 (20)	131.0394 (19)
Molinate	11.64	188.1109		126.0914 (100)	98.0973 (61)	83.086 (30)	272.0639 (26)	198.0328 (11)
Moxifloxacin[a]	5.93	402.1829		384.1696 (100)	402.1785 (40)	364.1646 (35)	260.0608 (25)	341.1542 (24)
Naproxen	10.84	231.1021		–	–	–	–	–
Nitrofurantoin	3.67	239.041		122.0111 (100)	95.0366 (44)	168.0409 (43)	100.0266 (39)	139.0133 (35)
Nitrofurazone	3.90	199.0467		199.0473 (100)	79.0186 (45)	108.0313 (32)	–	–
Norbenzoylecgonine	5.28	276.1236		136.0754 (100)	105.0327 (48)	154.0843 (35)	108.0807 (33)	68.0500 (14)
Norcocaine	5.87	290.1392		136.0762 (100)	168.1017 (51)	126.0118 (47)	108.0810 (27)	105.0336 (23)
Norfloxacin[a]	3.93	320.141		302.1302 (100)	233.1086 (56)	231.0551 (28)	205.0778 (25)	282.1245 (17)
Ochratoxin	11.65	404.0901		260.9924 (87)	239.0107 (59)	262.9923 (33)	221.0006 (32)	241.0102 (22)

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Ofloxacin[a]	3.78	362.1516		261.1044 (100)	221.0717 (11)	–	–	–
Olanzapine	4.84	313.1487		256.0886 (100)	198.025 (44)	213.0487 (35)	84.081 (31)	169.0762 (20)
Omeprazol	9.08	346.1225		180.0459 (100)	165.0245 (73)	136.0223 (38)	150.0920 (14)	–
Omethoate	2.92	214.0303		124.983 (100)	127.016 (74)	109.0051 (58)	78.9948 (20)	154.9929 (14)
Oxacillin	9.48	402.1123		265.0592 (71)	182.0253 (67)	160.0427 (46)	243.0769 (43)	114.0371 (31)
Oxadixyl	7.83	279.1345		132.0803 (100)	117.0578 (30)	–	–	–
Oxamyl	3.72	220.0756		72.0448 (100)	–	–	–	–
Oxolinic acid	6.74	262.0715		244.0595 (100)	216.0305 (30)	160.0394 (23)	172.0408 (11)	215.0209 (10)
Oxyfluorfen	12.82		360.0251	–	–	–	–	–
Oxymesterone	11.93	319.2273		113.0589 (99)	125.0590 (98)	319.2243 (61)	189.1645 (31)	
Oxytetracyclin[a]	4.83	461.156		426.1188 (100)	201.0550 (81)	283.0620 (56)	337.0716 (47)	365.0655 (38)
Paclobutrazol	11.51	294.1373		70.0406 (100)	125.0145 (21)	–	–	–
Pantoprazol	8.91	384.0829		200.0381 (100)*	138.0555 (30)*	–	–	–
Parathion-ethyl	12.54	292.0408		118.921 (100)	235.9797 (43)	94.0412 (37)	110.019 (34)	136.9317 (31)
Paroxetin	8.57	330.1505		70.066 (100)	192.1177 (98)	330.1493 (79)	123.0586 (45)	135.0606 (42)
Pefloxacin[a]	3.92	334.1567		316.1458 (100)	233.1079 (47)	205.0775 (21)	290.1615 (19)	203.0618 (12)
Pendimethalin	14.68	282.1454		212.0672 (100)*	194.0576 (14)*	–	–	–
Penicillin G	8.50	335.1065		198.0533 (70)	176.0708 (60)	160.0425 (53)	114.037 (50)	182.0256 (39)
Pipemidic acid[a]	3.30	304.1409		217.1086 (100)	189.0784 (65)	286.1322 (34)	215.0602 (24)	148.0514 (19)

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Piperacillin[a]	8.30	518.1709		398.0782 (100)	143.081 (97)	399.0799 (20)	160.0431 (20)	−
Pirimicarb[a,b]	9.15	239.1508		72.0456 (100)	182.1263 (29)	85.0762 (17)	109.0764 (17)	137.0713 (11)
Pravastatin	10.25	425.2539		327.1565 (41)		−	−	−
Promecarb	11.47	208.1337		109.0637 (100)	94.0417 (28)	−	−	−
Propanil ESI NEG	10.96		215.9983	159.9709 (100)	161.9689 (65)	215.9971 (21)	217.9961 (14)	163.9664 (11)
Propanil ESI POS	10.96	218.0139		127.0182 (100)	129.0160 (60)	161.9875 (43)	163.9851 (30)	200.0045 (12)
Propiconazole	13.15	342.0776		158.9762 (100)	160.9747 (66)	−	−	−
Propoxur	8.57	210.1130		93.0327 (100)	111.0431 (64)	65.0394 (399	148.9913 (28)	133.0138 (11)
Pymetrozine	3.50	218.1042		105.0443 (100)	−	−	−	−
Pyridaphenthion	11.75	341.0725		189.0649 (100)	205.0431 (42)	92.0487 (24)	363.0552 (23)	143.0613 (11)
Pyrifenox ISOMER 1	11.77	295.0405		93.0568 (100)	−	−	−	−
Pyrifenox ISOMER 2	12.07	295.0405		93.0571 (100)	−	−	−	−
Pyrimiphos-methyl[a,b]	13.21	306.1041		108.0547 (100)	164.1164 (68)	306.0977 (34)	136.0874 (30)	95.0609 (27)
Pyriproxifen[a,b]	14.46	322.1443		96.0443 (100)	185.0574 (24)	129.0700 (24)	119.0490 (19)	134.0726 (15)
Quizalofop-ethyl	14.05	373.0955		299.0589 (100)	255.0349 (62)	271.0647 (46)	301.0592 (35)	192.0699 (29)
Reserpine	9.42	609.2812		609.2802 (100)	195.0636 (22)	397.2115 (21)	174.0902 (16)	−
Risperidone	6.45	411.2196		191.1176 (100)	−	−	−	−
Roxythromycin[a]	10.57	837.5324		158.1187 (100)	679.4395 (57)	837.5307 (18)	116.1051 (10)	−
Simazine[a]	8.44	202.0859		96.0557	104.0012	124.0868	132.0320	68.0255

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
				(100)	(98)	(85)	(80)	(47)
Simvastatin	14.63	419.2797		441.2631 (100)	325.1760 (21)	–	–	–
Spinosyn A	12.44	732.4687		142.1236 (100)	732.4708 (37)	–	–	–
Spinosyn D	12.89	746.4843		142.1223 (100)	746.4833 (37)	–	–	–
Spiroxamine	10.65	298.2746		144.1384 (100)	100.1119 (67)	–	–	–
Stanozolol	12.61	329.2593		329.2582 (100)	–	–	–	–
Sulfadiazine[a]	2.87	251.0602		92.0490 (100)	108.0436 (75)	156.0096 (31)	96.0552 (25)	145.0768 (16)
Sulfamethazine[a]	4.51	279.0915		124.0863 (100)	92.0489 (79)	108.0441 (74)	186.0315 (48)	156.0106 (34)
Sulfamethiazol	4.14	271.0323		92.0500 (100)	108.0448 (80)	156.0124 (78)	271.0330 (49)	–
Sulfamethoxazol	4.88	254.0599		92.0492 (100)	108.0443 (78)	156.0090 (28)	254.0601 (26)	99.0548 (15)
T-2 toxin	11.08	467.2281		245.1172 (23)	387.1401 (15)	–	327.1202 (13)	–
Tebufenozide	12.54	353.2229		133.0653 (100)	225.1386 (28)	–	105.0702 (24)	–
Tebufenpyrad	14.28	334.1686		334.1698 (100)	145.0545 (80)	117.0225 (74)	336.1704 (33)	147.0512 (30)
Teflubenzuron	14.40		378.9664	195.9532 (100)	338.9534 (83)	197.9509 (66)	340.9522 (53)	294.9635 (39)
Terbacil	8.72		215.0588	158.9955 (100)	160.9905 (33)	–	–	–
Terbumetone[a,b]	10.34	226.1668		170.0965 (100)	100.0507 (41)	114.0664 (38)	142.0722 (34)	97.0399 (22)
Terbuthylazine[a]	11.28	230.1172		174.0537 (100)	176.0533 (35)	104.0012 (32)	132.033 (24)	79.0064 (23)
Terbutryn[a]	11.83	242.1439		186.0732 (100)	91.0325 (52)	96.0556 (41)	138.0779 (27)	116.0286 (27)
Tetraconazole	12.21	372.0293		158.9774 (100)	160.9745 (63)	70.0401 (19)	–	–

Compound	RT (min)	Exact mass[c] [M + H]+	Exact mass[c] [M-H]−	Product ion 1 m/z	Product ion 2 m/z	Product ion 3 m/z	Product ion 4 m/z	Product ion 5 m/z
Thiabendazol[a,b]	5.10	202.0439		175.0332 (100)	131.0607 (46)	202.0439 (27)	–	–
Thiacloprid	6.61	253.0314		126.0100 (100)	128.0082 (34)	–	–	–
Thiamethoxam	4.26	292.0271		131.9670 (100)	122.0710 (56)	152.0279 (45)	181.0543 (40)	133.9644 (36)
Thiobencarb	13.39	258.0719		125.0153 (100)	127.0129 (33)	–	–	–
Thiodicarb	9.36	355.0568		88.0214 (93)	113.0134 (11)	107.9936 (10)	–	–
Thiophanate-methyl	8.37	343.0534		151.0333 (100)	160.0518 (12)	85.9699 (11)	–	–
Tolyfluanid	12.86	346.9858		137.0301 (100)	237.9683 (100)*	137.0301 (86)*	239.9657 (65)*	368.9612 (60)*
Triadimenol	12.01	296.1166		318.1002 (100)*	296.1177 (55)*	227.0868 (46)*	99.0805 (40)*	320.0997 (32)*
Trichlorfon	5.83	256.9304		127.0168 (100)	109.0055 (71)	78.9954 (31)	93.0114 (22)	96.9615 (18)
Triflumizole	13.83	346.0934		278.0586 (100)	346.0946 (50)	178.9892 (36)	280.0561 (33)	200.0097 (28)
Trimethoprim[a]	3.78	291.1457		123.0664 (100)	261.0978 (80)	230.1137 (78)	275.1142 (58)	291.1381 (47)
Tylosin	9.32	916.5270		916.5278 (100)	174.1122 (55)	772.4484 (20)		–
Venlafaxin	6.71	278.2120		121.0647 (100)	147.0806 (44)	91.0545 (16)	173.0975 (14)	159.0811 (14)
Zearalonone	11.48	319.1545		185.0606 (29)	187.0769 (25)	203.0722 (21)	157.660)	–

*Important product ions in the LE function.

ªCompounds studied in section "Effect of lockspray parameters on mass accuracy", section **"**Effect of Enhanced Duty Cycle in sensitivity and mass accuracy", section "Optimization of chromatographic conditions" and section "Study of fragmentation conditions"

ᵇCompounds studied in the section "Effect of V and W ion optic modes on mass accuracy. Matrix effect"

[c]There is a difference of around 0.5 mDa between the theoretical mass and the calculated mass, as the software used adds or subtracts the mass of a hydrogen atom instead of the mass of a proton.

HPLC-grade water was obtained from deionized water passed through a Milli-Q water purification system (Millipore, Bedford, MA, USA). HPLC-grade methanol (MeOH) and acetonitrile (ACN) were purchased from ScharLab (Barcelona, Spain). Formic acid (HCOOH) (>98% w/w) was obtained from Fluka. Sodium hydroxide (>99%) was obtained from ScharLab. Leucine enkephalin was purchased from Sigma Aldrich.

Working standards of 2.5, 10, 25, 50, 100 and 250 μg/L were prepared from 50 mg/L mixed standards solutions in acetone, by diluting with water.

Instrumentation

An Acquity ultra-performance liquid chromatography (UPLC) system (Waters, Milford, MA, USA) was interfaced to a QTOF mass spectrometer (QTOF Premier, Waters Micromass, Manchester, UK) using an orthogonal Z-spray-electrospray interface. The LC separation was performed using two Acquity UPLC BEH C_{18} 1.7 μm particle size analytical columns of 50 × 2.1 mm and 150 × 2.1 mm (both from Waters), at a flow rate of 300 μL/min. The mobile phases used were A = H_2O and B = MeOH, both with 0.01% (v/v) HCOOH. The proportion of MeOH was linearly increased from 10% to 90% in 5 min (14 min for the 150 mm column), followed by a 1 min isocratic period (2 min for the 150 mm column) and returned to the initial conditions with a total run time of 8 min (18 min for the 150 mm column). The injection volumes were 20 and 50 μL for the 50 and 150 mm columns, respectively. Nitrogen (Praxair, Valencia, Spain) was used as both the drying gas and the nebulising gas. The gas flow rate was set at 600 L/h. The resolution of the TOF mass spectrometer was ~10 000 at full width half maximum (FWHM) in V-mode and 17 500 FWHM in W-mode, at *m/z* 556. MS data were acquired over an *m/z* range of 50–1000 in a scan time of 0.3 s. The MCP detector potential was set to 1750 V. A capillary voltage of 3.5 kV and a cone voltage of 25V in positive ionization mode were used. The collision gas was argon (99.995%, Praxair). The interface temperature was set to 350°C and the source temperature to 120°C. The column temperature was set to 40°C for the 50 mm column and 60°C for the 150 mm one (in order to minimize back-pressure).

For automated accurate mass measurement, the lock-spray probe was used, using as lockmass the [M + H]+ ion of leucine enkephalin (2 μg/mL) in ACN/water (50:50) at 0.1% HCOOH pumped at 30 μL/min through the lock-spray needle. A cone voltage of 60–70 V was selected and checked daily to obtain adequate signal intensity for this compound (~500 counts). The [M + H]+ ion of leucine enkephalin at *m/z* 556.2771 was used to recalibrating the mass axis and ensure a robust accurate mass measurement over time.

Calibration of the mass-axis was performed weekly using the built-in single-syringe pump, directly connected to the interface. Calibration from *m/z* 50 to 1000 was conducted with a 1:1 mixture of 0.05 M NaOH/5% (v/v) HCOOH diluted (1:25) with water/ACN (20:80 v/v) plus imazalil (yielding a [M + H]+ ion at *m/z* 297.0561) at a final concentration of 500 μg/L.

The data station operating software was Masslynx version 4.1 (Waters).

RESULTS AND DISCUSSION

As stated in the introduction, LC/MS still does not have commercially available mass spectral libraries due to the non-availability of a standardized interface. This implies the building of home-made libraries to facilitate the searching, with a focus on particularly relevant compounds.

Building an empirical library database

Several experiments were carried out to evaluate and optimize the main parameters that can affect the sensitivity, mass accuracy and fragmentation provided by UHPLC coupled to TOF MS. The objective of our work was to establish the optimum working conditions in order to obtain reliable spectra to be included in the library for further non-target analysis. To this aim, a significant number of compounds (43, including pesticides and antibiotics), with *m/z* values ranging from 192 ([M + H]+ ion of carbendazim) to 837 ([M + H]+ ion of roxythromycin), were selected to provide more complete and realistic knowledge of the instrument potential.

In all these experiments, 20 mDa narrow-window extracted ion chromatograms (nwXICs) were generated for each analyte at the *m/z* value of its [M + H]+ ion, and the mass spectra were combined over the corresponding chromatographic peaks. Mass accuracy was calculated as the difference between the theoretical exact mass and the experimental one, and expressed in mDa units.

Effect of the lockspray parameters on the mass accuracy

The QTOF Premier mass spectrometer is equipped with a lockspray interface, independent of the sample interface, used for the introduction of the reference substance. This system avoids matrix suppression of the lock mass but it can affect the chromatographic peak shape if sampled when the top of the peak is eluting. The effect on the mass accuracy of the acquisition interval and the scan time acquisition of the lock mass was evaluated, trying to minimize the possible interference of the lock mass in the analyte peak. Default settings given by the software were 1 s scan time and 10 s acquisition interval. To perform this experiment, a 50 µg/L mix standard was analyzed at different lockspray scan times (0.25, 0.5 and 1 s) and acquisition intervals (10, 20, 30 and 50 s).

As shown in Table 2, despite increasing the acquisition interval of the lock mass as well as decreasing the scan time, the mass accuracy was not affected very much. In fact, after one-way analysis of variance (ANOVA) for each compound studied, we did not find significant differences in the mass error when changing lockspray parameters.

Table 2. Effect of the frequency (s) and scan time (s) of lock mass acquisition in mass accuracy

Interval (s) at 1 s scan time	Mean error ± SD (mDa)	Scan time (s) at 10 s interval	Mean error ± SD (mDa)
10	0.63 ± 0.45	1.0	0.63 ± 0.45
20	0.67 ± 0.44	0.5	0.65 ± 0.57
30	0.64 ± 0.52	0.25	0.54 ± 0.50
50	0.49 ± 0.41	0.1	0.60 ± 0.41

We also tested different combinations of these parameters to evaluate their joint effect on the mass accuracy measurements. Thus, when an interval of 30 s or higher was used together with low scan times (less than 1 s), higher mass errors were obtained. For example, using scan times of 0.5 s and 0.25 s, and an interval of 50 s, average mass errors were 1.38 and 1.47 mDa, respectively, i.e., two-fold higher than those given in Table 2. In contrast, when an interval of 30 s or higher was selected, the number of peaks interfered by the lock mass acquisition was minimized (less than 4% of the compounds studied were affected at the peak top). Finally, the

best option for lock mass conditions, for both adequate peak shape and mass accuracy, was an interval of 30 s and a scan time of 1 s when using UHPLC.

Another parameter studied was the influence of the signal intensity of the lock mass (measured as counts per second, cps) on the mass accuracy. For this purpose, the 50 µg/L mix standard was analyzed at 300, 500 and 800 cps for lock mass (these intensities were adjusted by varying the cone voltage). It was concluded that the signal intensity of the lock mass did not significantly affect the mass accuracy of analytes, although a slight decrease in mass error was observed in the range 300–500 cps (data not shown). Finally, a lock mass signal intensity of between 400 and 500 cps was selected as the optimum value.

Effect of the ion abundance on the mass accuracy. Programmable dynamic range enhancement

Another parameter studied was the relationship between mass accuracy and analyte ion abundance. Traditional TDC detectors suffer from saturation effects, which limit some applications of the TOF mass spectrometer. In our instrument, a TWIG is used to extend the dynamic range of the TDC detector. The entrance lens to the TWIG is used as a gate to allow only a portion of ions to pass through, therefore producing an attenuated pulsed signal. The TWIG electrodes are then used to smooth out the attenuated signal and to create a homogeneous ion beam of lower intensity.

To assess the dynamic range increase when using the TWIG collision cell, we compared the data obtained when acquiring with, and without, the attenuated pulsed signal (DRE) enabled. The mass accuracy, over a wide dynamic range, was also compared. Differences between the measured and theoretical accurate masses for 43 compounds at diverse ion abundances (i.e. different concentrations) were studied. In both cases (DRE-off and DRE-on), similar mass accuracy was obtained when the ion intensity was sufficiently small not to saturate the detector (less than approximately 3500 counts). However, when the ion intensity was higher than 3500 counts, pDRE-enabled acquisitions demonstrated a significant improvement in the mass accuracy compared with the acquisitions without pDRE. As an example, Fig. 1(a) shows the effect of DRE on average mass accuracy for six of the compounds selected (carbendazim, thiabendazol, imazalil, simazine, terbuthylazine and pirimiphos-methyl) at different concentration levels. Selection of the representative compounds was performed as a function of the sensitivity

in order to obtain similar saturation for the concentration levels tested. Our results demonstrated that mass accuracy was satisfactory using the pDRE option, with mass errors lower than 1 mDa at all concentrations tested, illustrating the high degree of confidence, with regard to mass accuracy, provided by the instrument.
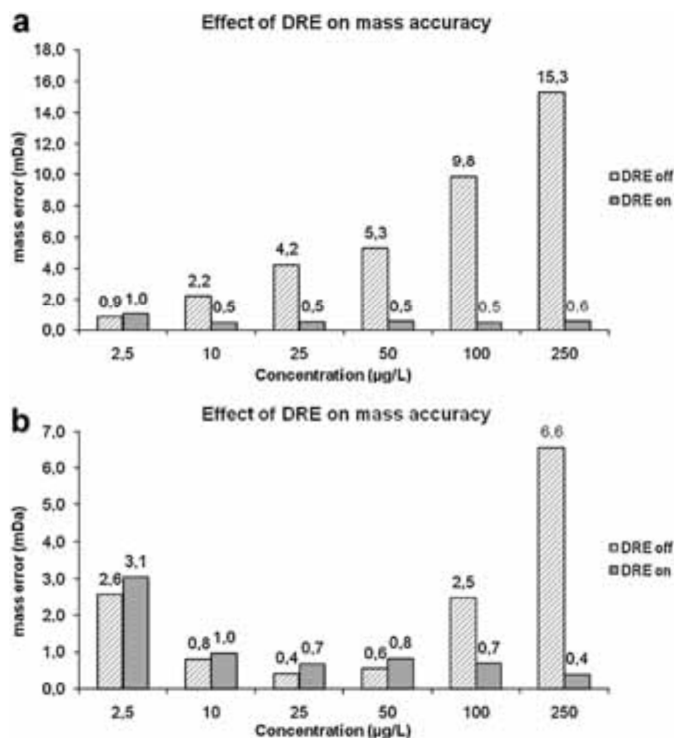
Figure 1. Effect of DRE on average mass accuracy for several pesticides at different concentrations (n = 3). (a) Compounds showing good sensitivity (good ion statistics): carbendazim, thiabendazol, imazalil, simazine, terbuthilazine and pirimiphos-methyl. (b) Compounds showing poor sensitivity (poor ion statistics): alachlor, dimethoate, metolachlor, pyridaphention, pyriproxyfen and thiobencarb.

Consequently, pDRE can be satisfactorily employed to increase the ion abundance range for which accurate mass measurements are recorded. Although the mass errors achievable were comparable when the ion intensities were lower than 3500 counts, the mass accuracy was notably improved at high concentrations (high ion abundance) when using this option. This significant improvement illustrates its effectiveness in qualitative analysis. It is important

to note that applying this approach affects the number of spectra per second recorded, decreasing the number of points per chromatographic peak. However, this was not a problem, even when using UHPLC, because of the high acquisition speed of the instrument.

On the other hand, when the ion statistics are poor, the mass accuracy is also affected in the TDC detector even when using the pDRE option. Figure 1(b) shows this effect for several compounds that presented low sensitivity (alachlor, dimethoate, metolachlor, pyridaphention, pyriproxyfen and thiobencarb), which were injected at different concentrations. The results illustrate the effect of the ion abundance on the mass accuracy. However, even in this adverse case, the mass errors are still acceptable at analyte concentrations down to 10 µg/L.

Effect of the enhanced duty cycle in the sensitivity and the mass accuracy

The sensitivity of TOF instruments is a key parameter in environmental and food safety analysis. The TWIG transports the ions in packets and so it is possible to synchronize the TOF pusher with individual ion packet and enhance the duty cycle over a selected *m/z* range.[22] This mode of operation is referred as enhanced duty cycle (EDC).

A 2.5 µg/L mix standard of 43 pesticides and antibiotics (see Table 1) was selected as a compromise for detecting the majority of the analytes but avoiding saturation. The mix standard was injected with and without EDC mode at *m/z* 200, 300, 400, 500 and 800. The objective was to evaluate the effective increase of ion abundance in the mass range between *m/z* 200 and 800, as well as the mass interval affected. As can be seen in Fig. 2, the EDC increased the analyte sensitivity between 2 and 6 times in relation to the measurements without EDC.
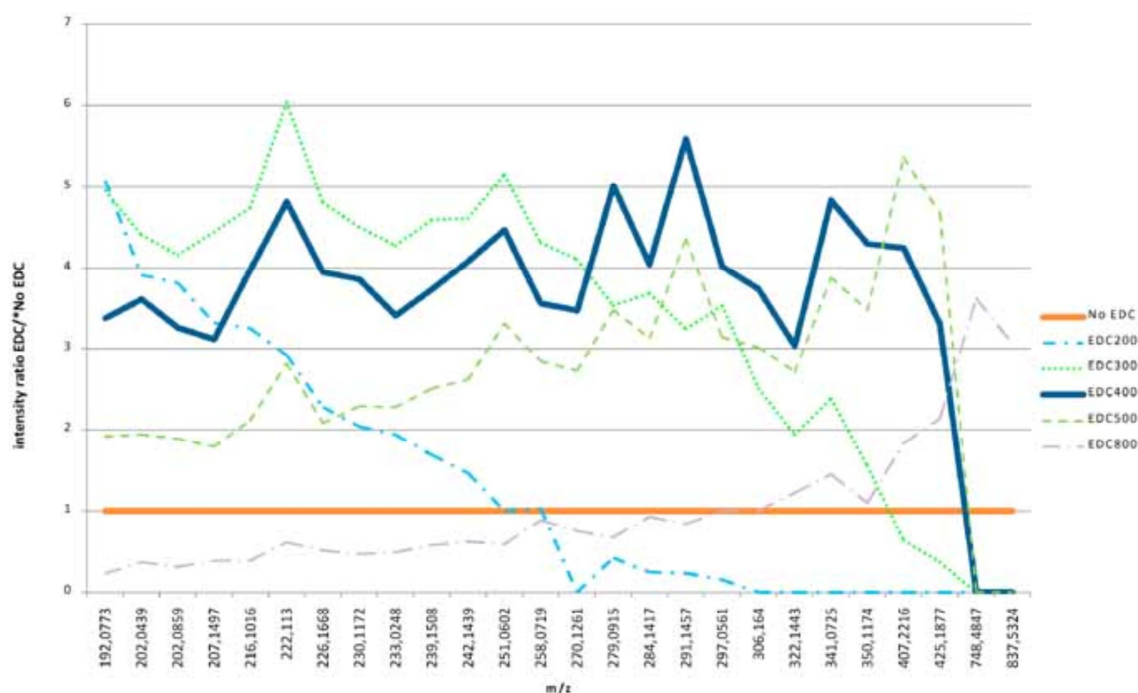
Figure 2. EDC/No EDC ion abundance ratio at different *m/z* values for organic pollutants with *m/z* ranging from 192 to 838. It shows the EDC effect on TOF MS signal.

The mass range affected was found to depend on the target mass range selected. Thus, when EDC values at *m/z* 200 or 800 were selected (i.e. located at both ends of the mass spectra), the sensitivity increased only for analytes in a range of approximately 100 *m/z* units around the *m/z* value tested. However, when EDC values at *m/z* 300, 400 or 500 were selected, the sensitivity in all the mass range was improved. Thus, to detect and identify target analytes and/or elucidate unknowns, the most useful approach to increase the sensitivity would be to use the EDC function at *m/z* 400, as the sensitivity for ions up to *m/z* 700 would notably increase. We also observed that the repeatability of the ion intensity measurement was not affected by the use of the EDC.

The possible influence of the EDC on the mass accuracy was also investigated. Although a direct relationship was not expected between these two parameters, our data indicated some kind of correlation because a slight loss of mass accuracy was observed when the EDC function was selected. The mass error was around two-fold higher when the EDC function was switched on, but it was lower for those analytes with ions at *m/z* values close to the selected EDC value, as shown in Table 3.

Table 3. Average mass error (mDa) for 30 selected compounds as a function of the mass range and EDC applied

Mass range (m/z)	No EDC	EDC m/z 200	EDC m/z 300	EDC m/z 400	EDC m/z 500	EDC m/z 800
<250	0.7	0.8	2.1	2.5	2.3	2.4
250-350	0.5	2.8	1.0	1.3	1.6	1.4
350-450	0.6	–	3.3	1.0	1.1	1.3

(-) For those analytes with m/z higher than 350, no peaks were detected or the sensitivity was too low to correctly calculate the mass error.

Finally, the EDC was not selected for the screening of compounds that might be present in samples in a wide range of concentrations. As the EDC cannot be applied simultaneously with the pDRE and because of the possibility of saturation occurring for the most abundant compounds, the latter option (pDRE) was preferred.

Effect of calibration on the mass accuracy

When the data described in the section '*Effect of the lockspray parameters on the mass accuracy*' were studied in detail, a remarkable increase in the mass errors was observed for those compounds with [M + H]+ ions ranging from m/z 270 to 310. After considering possible reasons, we noticed that the highest mass error in the calibration function (see Fig. 3(a)) was also produced in this interval. For mass axis calibration of the TOF mass spectrometer, adducts of sodium formate (obtained from NaOH 0.05M/HCOOH 5% (v/v) diluted 1:25 in water/ACN (20:80)) have traditionally been used in our laboratory. From the calibration report using this solution, we observed that the ion at m/z 294 had a notably lower intensity. This led to higher mass deviations in the area around m/z 300 that could not be corrected by the calibration function of fifth polynomial order, therefore affecting the accurate mass measurements.
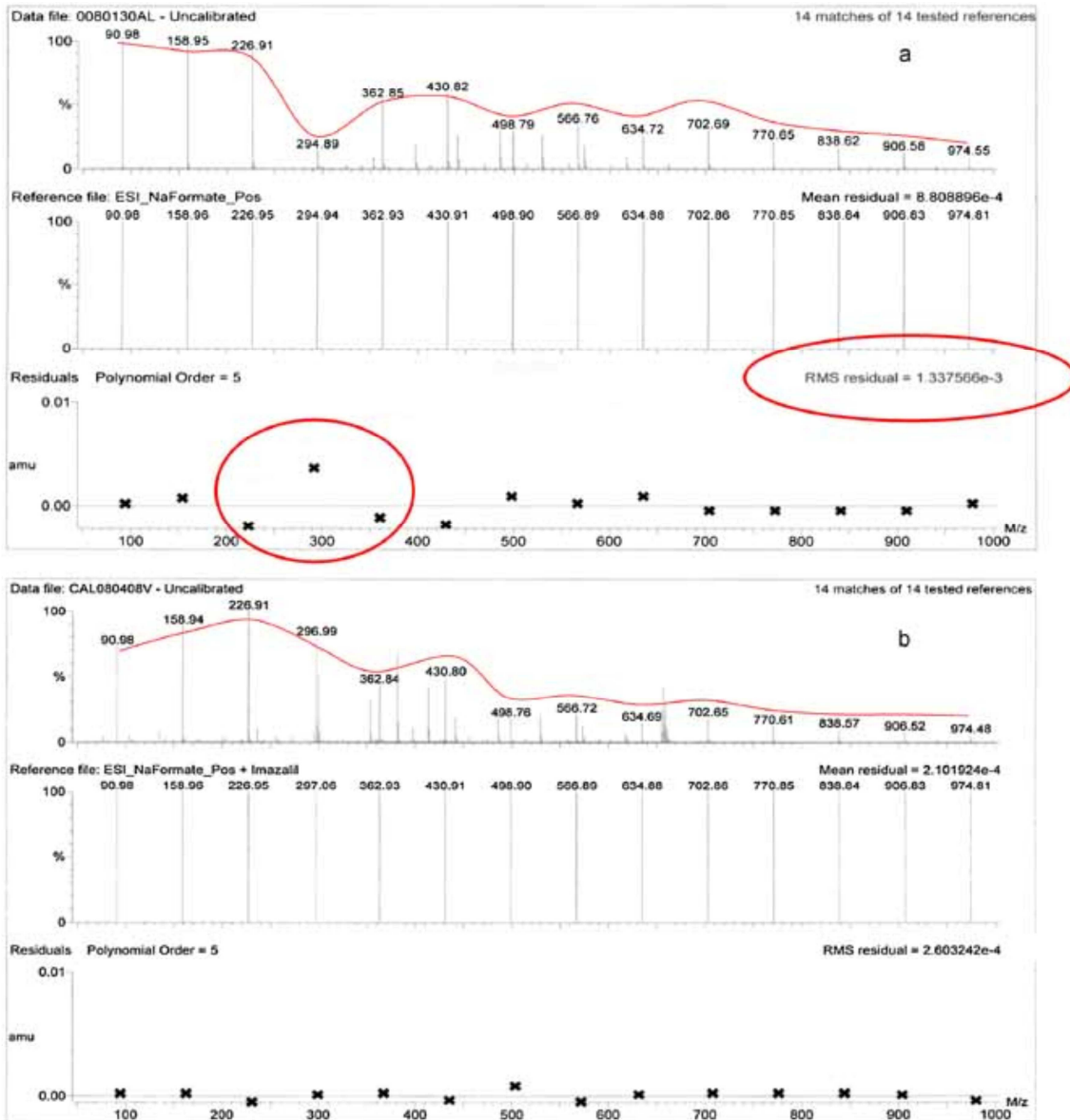
Figure 3. Calibration report obtained with (a) sodium formate and (b) sodium formate + imazalil. Residuals for *m/z* range 226 to 362 are highlighted.

To solve this problem, different compounds with $[M+H]^+$ ions around m/z 300 were studied. The fungicide imazalil (m/z 297.0561) was tested at different concentration levels (5, 50 and 500 µg/L) in the calibration solution, and 500 µg/L was finally selected as the best option. In this way, the peak envelope was smoothed and mass errors were lower (Fig. 3(b)). As illustrative examples, pesticides metolachlor (m/z 284.1417) and buprofezin (m/z 306.1640) had mass errors of 0.7 and 1.5 mDa, respectively, when only sodium formate adducts were used. However, after adding imazalil to the calibration solution, the mass errors decreased to 0.2 and 0.7 mDa, respectively.

Effect of the V and W ion optic modes on the mass accuracy. Matrix effect

Two mass spectrometry acquisition methods – selecting V and W ion optics modes – were compared. A separate calibration file was recorded for each mode. The effect of using V-mode (resolution 10 000 FWHM) or W-mode (resolution 17 500 FWHM) on mass accuracy was evaluated for five different matrices: two food matrices (cucumber, pepper) and three waters (surface water, influent and effluent urban wastewater). Samples were spiked with 10 organic contaminants (see Table 1, compounds marked as 'b') with $[M+H]^+$ ions ranging from m/z 192 to 350, at 50 µg/L level each and analyzed in triplicate. As shown in Table 4, slightly lower mass errors were typically obtained when the W-mode was used. As an example, average errors for influent wastewater (the most complex water matrix) decreased from 1.30 mDa (V-mode) to 0.81 mDa (W-mode).

Table 4. Average mass errors for 10 selected compounds ($n=3$) analyzed in the five food and environmental matrices, in both V and W ion optics acquisition modes

	Mean mass errors (mDa) ($n=3$)					
	Solvent	**Cucumber**	**Pepper**	**EWW**	**IWW**	**SW**
V-mode	1.09	1.15	1.25	1.33	1.30	1.41
W-mode	0.78	1.03	1.30	0.99	0.81	0.93

EWW = effluent wastewater; IWW = influent wastewater; SW = surface water

However, a significant decrease in sensitivity was also observed (up to 3-fold in some cases) when using the W-mode. In food matrices with important matrix suppression effects, such as pepper, the mass errors were even higher due to the sensitivity decrease. Thus, the W-mode

was considered as a good approach for the analysis of complex matrices in those cases when sufficient sensitivity was maintained.

Optimization of the chromatographic conditions

Another alternative way to reduce the mass errors in analysis of complex matrices would be the use of longer chromatographic columns. All the experiments described until now were carried out with a 50 mm UHPLC column. At this point, we tested the effect of using a three-fold longer column, i.e. 150 mm long. In this way, better separation between analytes and interferents, and consequently better mass accuracy, was expected. Using the 150 mm column and working in V-mode, the mass errors were similar to those in W-mode for all the matrices tested, but without sensitivity loss. Using a longer column, it was also possible to increase the injection volume. We tested a 50 µL loop, instead of the 20 µL loop initially used, obtaining excellent peak shape while injecting 2.5 times more sample. As a consequence, the sensitivity could be increased by a factor of around 2.5 times although with the disadvantage of longer chromatographic runs. Finally, a 150 mm UHPLC column (Acquity UPLC BEH C18) was selected for the empirical spectra library creation and for further screening experiments. For the organic mobile phase, we tested the use of MeOH and ACN, as well as the addition of HCOOH. For most of compounds studied, the sensitivity was better when using MeOH with 0.01% (v/v) HCOOH (see Instrumentation section).

Study of the fragmentation conditions

To improve confidence in positive findings, analyte fragmentation was promoted, thus producing richer mass spectra, which are more useful for identification/elucidation purposes. The objective was to find a compromise between producing enough fragment ions and still observing the $[M + H]^+$ ion. Different approaches were considered using both in-source and collision cell fragmentation.

Different cone voltages were tested to evaluate sensitivity (10, 20 and 30 V) as well as in-source fragmentation (40, 50 and 60 V). The sensitivity was maximized at cone voltages ranging from 20 to 30 V. However, it was unfeasible to find a common voltage for the pesticides and antibiotics studied at which fragmentation was acceptable for all compounds. Thus, two different cone voltage ramps, 10–70 V and 20–60 V, were tested in an attempt to find compromise fragmentation conditions. The empirical results revealed a considerable loss of mass accuracy, with mass errors mostly exceeding the value of 5 ppm normally required for a correct

identification, even with increasing the scan time (Fig. 4). In addition, the cone voltage ramps did not produce fragmentation as efficiently as a single high voltage, probably because the algorithm of the ramp is not linear.
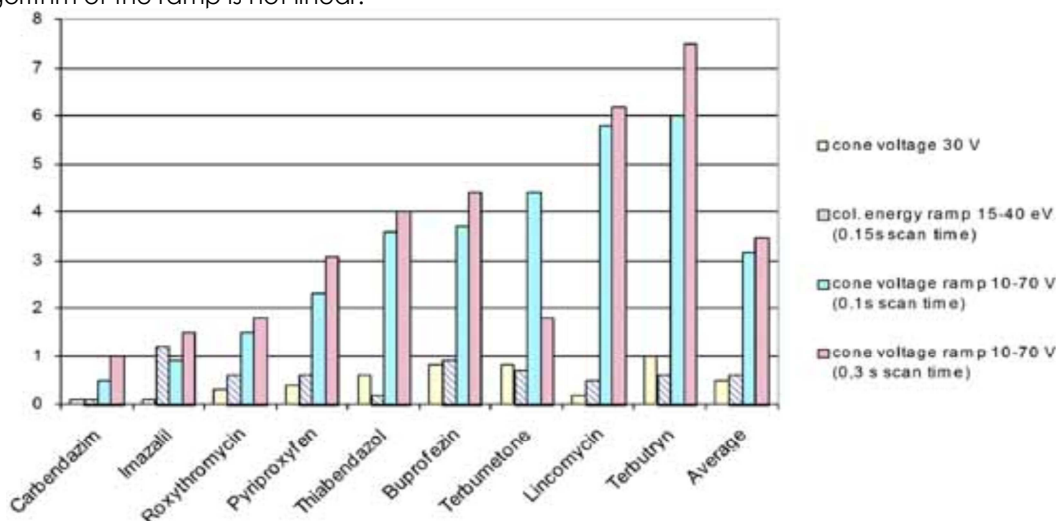
Figure 4. Mass errors (mDa) of the base peak of the spectra using a cone voltage ramp of 10–70 V at two different scan times, and a collision energy ramp 15–40 eV. Comparison with the non-promoted fragmentation acquisition, at fixed cone voltage 30 V.

To improve the identification potential of (Q)TOF MS, collision cell fragmentation was evaluated. As stated before, our instrument is equipped with a TWIG collision cell, designed to always work with an argon gas flow-through. This permits the entrance energy in the TWIG to be increased, thus promoting collision-induced dissociation (CID), although working in TOF MS mode. After testing the possibility of using single collision energy (CE) to fragment the analytes (no satisfactory compromise CE was found), four CE ramps were tested: 10–30 eV, 10–50 eV, 10–70 eV, and 15–40 eV. A CE ramp of 15–40 eV was finally chosen as it involved a smaller voltage change, with improved fragmentation. In addition, collision cell fragmentation revealed low errors for the main product ions, typically below 1 mDa, lower than those obtained after in-source fragmentation (see Fig. 4).

Due to better fragmentation and lower mass errors, two simultaneous TOF MS acquisition functions were finally monitored: the first one with 4 eV collision energy (low energy function, LE, poor or no fragmentation) and the second one with a collision energy ramp ranging from 15 to 40 eV (high energy function, HE, promoting collision cell fragmentation). The LE and HE function

scan times were set to 0.2 s and 0.15 s, respectively, with an interscan delay of 0.05 s. This type of acquisition is known as MS^E. Furthermore, the spectra obtained from the HE and MS/MS acquisitions were quite similar in most cases (see Fig. 5). Moreover, additional information such as isotopic pattern and adduct formation is retained (normally not available in MS/MS experiments, where unit mass resolution is commonly selected in the first quadrupole mass analyzer). For example, the chlorine pattern for the fungicide imazalil ([M + H]$^+$, *m/z* 297) is also shown for the product ions at *m/z* 158, 176 and 255, demonstrating that this approach is excellent for identification purposes.[29]

Finally, for the creation of the in-house empirical spectral library, different standard pollutant solutions (see Table 1) at approximately 100 µg/L were injected into the UHPLC-(Q)TOF MS system under the conditions previously established, i.e. pDRE-on, V ion optics mode using a lock mass intensity around 500 cps with a 1 s scan time and 30 s acquisition interval, and a 150 mm UPLC column. In total, 230 compounds were included in the empirical library to facilitate screening in water and food samples. Different families of pesticides, antibiotics, pharmaceuticals, drugs of abuse and mycotoxins commonly detected were included in this list, as well as some metabolites, degradates and other transformation products. Compounds were acquired in either positive or negative ion electrospray mode, except for propanil, which was acquired at both polarities.

For each compound two library entries were created, one for HE and the other for LE acquisition yielding 462 entries in the empirical mass spectral library. Entries included compound name, nominal mass, exact mass and molecular formula. Any information field could be used to filter the 'hit list' and facilitate assignment of the right candidate. In Table 1, accurate masses for up to five abundant product ions (with their respective normalized intensity) are reported.
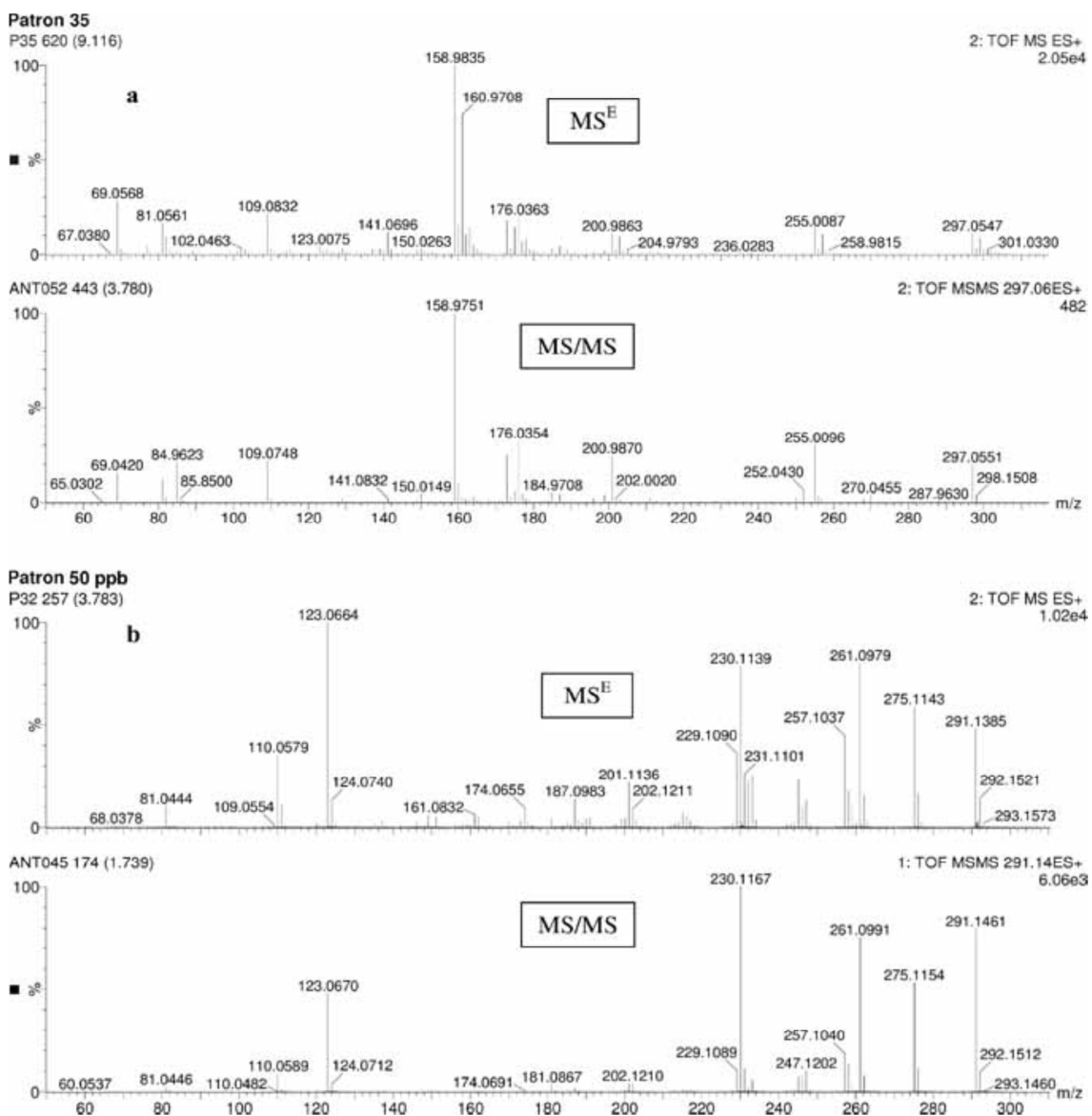
Figure 5. Comparison of MS spectra for (a) protonated fungicide imazalil and (b) protonated antibiotic trimethoprim reference standards by using MS^E(top) and MS/MS (bottom).

Evaluation of the empirical library for screening purposes

Once the library was created, it was tested for the analysis of real-world samples. For this test an effluent wastewater sample was 50-fold pre-concentrated by passing 50 mL of sample through an Oasis HLB 60 mg solid-phase extraction (SPE) cartridge. The 5mL-eluted MeOH extract

was evaporated to dryness and reconstituted with 1mL H$_2$O/MeOH (90:10). Then 50 μL of the final extract were injected into the UPLC-(Q)TOF MS system under the chromatographic conditions explained in the paper. As shown in Fig. 6, the TIC was too complex to detect the trace level compounds; so data processing was required. We used ChromaLynx software in non-target mode for TIC deconvolution in the LE function, thus facilitating the detection of trace level components present in the sample. To illustrate the strategy applied, Fig. 6 shows the detection and identification of the fungicide carbendazim in effluent wastewater. The LE spectrum of the compound given by the software was searched against the in-house mass spectral library and carbendazim was proposed as the first candidate. In addition, the higher fragmentation observed in the HE spectrum allowed us to reliably improve the identification process.

At present, this non-target approach using an empirical spectral library is being compared with the target approach. For the target approach, a large list of compounds is specifically searched for in the full-scan TOF spectra after MS data acquisition (i.e. post-target) in different types of water and food matrices.
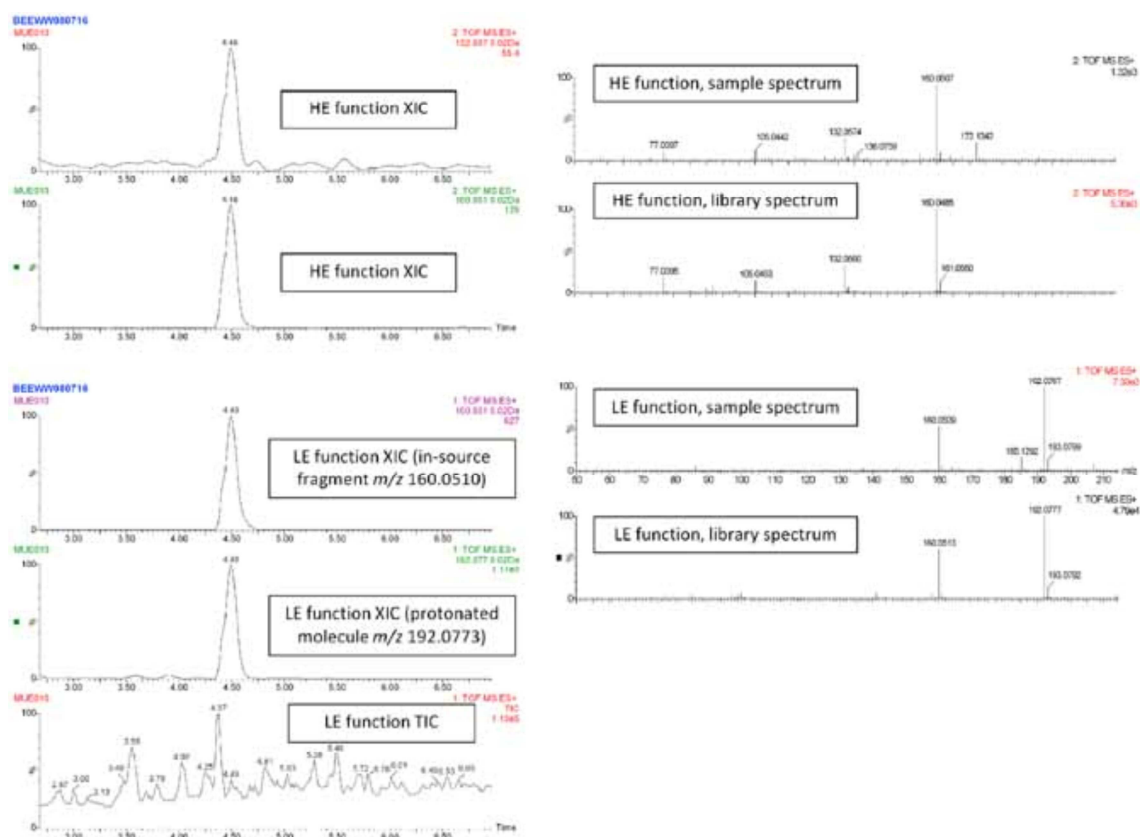
Figure 6. Effluent wastewater sample positive to carbendazim. Left: Total ion and extracted ion chromatograms in LE and HE functions for the EWW sample. Right: Comparison of carbendazim library spectra and suspected compound spectra in LE and HE functions.

CONCLUSIONS

An extensive investigation of the different factors that can affect mass accuracy, sensitivity and dynamic range in a UHPLC/QTOF MS instrument has been carried out. In addition, conditions for building a mass spectral library have been studied with the aim of improving the sensitivity and the analyte fragmentation information.

These results have allowed us to improve our knowledge of the capabilities, limitations and robustness of the instrument. This was necessary to enable us to perform an efficient wide-scope screening analysis at a later stage. Establishing the optimum experimental conditions to

build an empirical in-house spectral library is essential to optimize accuracy and future results of screening. The sensitivity and dynamic range of the instrument are now well-known, and this is crucial when dealing with the elucidation of unknown compounds. Under optimum conditions, mass errors below 1 mDa have been obtained for the standard solutions of the compounds investigated, and below 2 mDa for several food and water spiked samples. Good mass accuracy and sensitivity have been shown over a concentration range of three orders of magnitude when using the pDRE option on the Q-TOF Premier mass spectrometer.

An in-house compound spectral library has been created (and the main fragment ions reported) for 230 compounds, including pesticides, antibiotics, other pharmaceuticals, drugs of abuse and toxins, as well as several transformation products. Using this empirical library allows the rapid screening of samples making use of the MS^E approach (simultaneous acquisition at low and high collision energy), as the detection and the confirmation of the identity of the analyte can be achieved simultaneously, without the need for sample re-analysis.

ACKNOWLEDGEMENTS

REFERENCES

1. J. R. Kenseth, S. J. Coldiron. *Curr. Opin. Chem. Biol.* 2004, **8**, 418.

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

3. E. J. Want, B. F. Cravatt, G. Siuzdak. *ChemBioChem.* 2005, **6**, 1941.

4. S. Grimalt, O. J. Pozo, J. V. Sancho, F. Herna´ndez. *Anal. Chem.* 2007, **79**, 2833.

5. F. Herna´ndez, J. V. Sancho, M. Iba´n˜ez, S. Grimalt. *TrAC Trends Anal. Chem.* 2008, **27**, 862.

6. F. Hernández, S. Grimalt, O. J. Pozo, J. V. Sancho. *J. Sep. Sci.* 2009, **32**, 2245.

7. M. E. Thurman, I. Ferrer, O. Malato, A. R. Ferna´ndez-Alba. *Food Addit. Contam.* 2006, **23**, 1169.

8. Y. Picó, M. Farré, C. Soler, D. Barcel.ó *J. Chromatogr. A* 2007, **1176**, 123.

9. M. Mezcua, O. Malato, J. F. García-Reyes, A. Molina-Díaz, A. R. Ferna´ndez-Alba. *Anal. Chem.* 2009, **81**, 913.

10. A. A. M. Stolker, P. Rutgers, E. Oosterink, J. J. P. Lasaroms, R. J. B. Peters, J. A. Van Rhijn, M. W. F. Nielen. *Anal. Bioanal. Chem.* 2008, **391**, 2309.

11. A. Kaufmann, P. Butcher, K. Maden, M. Widmer. *J. Chromatogr. A* 2008, **1194**, 66.

12. S. González, M. Petrović, M. Radetic, P. Jovancic, V. Ilic, D. Barcel.ó *Rapid Commun. Mass Spectrom.* 2008, **22**, 1445.

13. M. Petrović, D. Barcel.ó *J. Mass Spectrom.* 2006, **41**, 1259.

14. J. V. Sancho, Ó. J. Pozo, M. Ibáñez, F. Hernández. *Anal. Bioanal. Chem.* 2006, **386**, 987.

15. M. Iba´n˜ez, J. V. Sancho, O. J. Pozo, W. Niessen, F. Herna´ndez. *Rapid Commun. Mass Spectrom.* 2005, **19**, 169.

16. M. Ibáñez, J. V. Sancho, Ó. J. Pozo, F. Hernández. *Anal. Chem.* 2004, **76**, 1328.

17. M. Krauss, H. Singer, J. Hollender. *Anal. Bioanal. Chem.* 2010, , **397**, 943.

18. T. Bristow, J. Constantine, M. Harrison, F. Cavoit. *Rapid Commun. Mass Spectrom.* 2008, **22**, 1213.

19. I. V. Chernushevich, A. V. Loboda, B. A. Thomson. *J. Mass Spectrom.* 2001, **36**, 849.

20. I. Ferrer, E. M. Thurman. Liquid Chromatography Time-of-Flight Mass spectrometry. Principles, Tools and Applications for Accurate Mass Analysis, John Wiley, New York, 2009, 261.

21. P. J. Weaver, A.-M. Laures, J.-C. Wolff. *Rapid Commun. Mass Spectrom.* 2007, **21**, 2415.

22. M. Farré, M. Gros, B. Hernández, M. Petrovic, P. Hancock, D. Barcel.ó *Rapid Commun. Mass Spectrom.* 2008, **22**, 41.

23. J.-C. Wolff, C. Eckers, A. B. Sage, K. Giles, R. Bateman. *Anal. Chem.* 2001, **73**, 2605.

24. F. Calbiani, M. Careri, L. Elviri, A. Mangia, I. Zagnoni. *J. Mass Spectrom.* 2006, **41**, 289.

25. J.-C. Wolff, T. R. Fuentes, J. Taylor. *Rapid Commun. Mass Spectrom.* 2003, **17**, 1216.

26. A.-M. Laures, J.-C. Wolff, C. Eckers, P. J. Borman, M. J. Chatfield. *Rapid Commun. Mass Spectrom.* 2007, **21**, 529.

27. F. Hernández, O. J. Pozo, J. V. Sancho, F. J. López, J. M. Marín, M. Ibáñez. *TrAC Trends Anal. Chem.* 2005, **24**, 596.

28. M. Ibáñez, J. V. Sancho, F. Hernández, D. McMillan, R. Rao. *TrAC Trends Anal. Chem.* 2008, **27**, 481.

29. F. Hernández, L. Bijlsma, J. V. Sancho, R. Díaz, M. Ibáñez. *Anal. Chim. Acta* 2010, under revision.

2.2.1. Discusión de resultados

En análisis non-target resulta de enorme importancia poder disponer de bases de datos o librerías de espectros de contaminantes y TPS. En el caso de la cromatografía de gases acoplada a espectrometría de masas con ionización electrónica (GC-EI-MS) existen extensas librerías de espectros que facilitan la elucidación e identificación de los llamados unknown-known compounds, es decir aquellos compuestos desconocidos para el analista pero conocidos o reportados en la literatura. Este es el caso de librerías como la NIST o la Wiley, con más de dos cientos mil espectros disponibles que maximizan el porcentaje de éxito en la elucidación de contaminantes mediante análisis por GC-EI-MS.

Un caso muy diferente es el que ocurre con el acoplamiento LC-MS. Esta técnica está cada vez más extendida debido a sus excelentes propiedades en el análisis de compuestos polares y semipolares, entre las que se incluyen su elevada sensibilidad y selectividad, el mínimo tratamiento de muestra requerido normalmente y sus interfases de ionización débil que permiten determinar con facilidad la masa del compuesto en cuestión. Sin embargo, este último atributo conlleva que la eficiencia de los procesos de ionización sea muy variable en función de parámetros como la fase móvil utilizada en la separación, el voltaje de capilar o de cono aplicados, e incluso del diseño de la propia interfase de ionización. Esto hace complicada la creación de librerías de espectros LC-MS estandarizadas que puedan ser utilizadas de forma fiable para la identificación de compuestos desconocidos.

Recientemente han surgido algunas librerías para LC-MS, como por ejemplo MassBank o Metlin, que intentan solventar el problema de las condiciones de adquisición en la interfase mediante la inserción de diversas entradas para cada uno de los compuestos indicando el voltaje de cono, el tipo de interfase (ESI o APCI principalmente), el tipo de analizador y, en algunos casos, la casa comercial del instrumento con el que se adquirió dicho espectro. De este modo, se incrementa notablemente la posibilidad de encontrar un espectro similar al adquirido en el propio instrumento bajo condiciones similares, pero no asegura que las relaciones de

intensidad de los diferentes iones se cumplan, limitando la utilidad de estas librerías. Además, estas bases de datos de espectros suelen disponer de la información de masa en nominal y están limitadas a unos pocos compuestos, como fármacos y drogas en el caso de MassBank o metabolitos (generalmente humanos) en el caso de Metlin. Así pues, aunque resultan de utilidad, las librerías de espectros de masa disponibles para compuestos analizables mediante LC-API-MS tienen un uso restringido.

La alternativa es, por tanto, la creación de una base de datos o librería doméstica (home-made) en base a condiciones de análisis claramente establecidas en un instrumento concreto. Esto es el objetivo perseguido en en el artículo titulado "Building a mass specra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry", pues las condiciones óptimas seleccionadas para la creación de la librería serán la base de futuros análisis y estrategias ya que, como se ha mencionado anteriormente, variaciones en ciertas variables implican cambios que pueden llegar a ser muy importantes en los espectros de masa resultantes y en los tiempos de retención de los compuestos.

Lo que se pretendió en este artículo fue establecer unas condiciones experimentales para el screening lo más genéricas posibles y aptas para un amplio rango de analitos. Obviamente, fué necesario alcanzar un compromiso, pues la mayoría de variables estudiadas dependen en gran medida del analito en cuestión y algunas de ellas son incluso incompatibles. Este último es el caso de las opciones pDRE y EDC. La primera de ellas permite aumentar el rango lineal de concentración evitando en gran medida la saturación y, por tanto, la medida errónea de intensidad o área y de masa exacta. La segunda aumenta la relación S/N para un determinado y prefijado rango de valores de m/z, pero disminuye de forma notable para el resto de masas. Por este motivo, la opción pDRE fue seleccionada para el screening "universal", descartando así el EDC, válido únicamente para un limitado rango de masas.

Por otro lado, se compararon los modos V y W del analizador TOF mediante el análisis de muestras seleccionadas, con el fin de comprobar el efecto de matrices complejas sobre la masa exacta en función de la resolución del analizador de masas. Se comprobó una ligera mejora en la exactitud de masa en todas las matrices

exceptuando la más compleja, el pimiento, que se vió afectada por una disminución de la señal en modo W, de hasta tres veces. Se propuso mejorar la separación cromatográfica, trabajando en modo V, triplicando la longitud de la columna, como alternativa al aumento de la resolución de masa, obteniendo valores muy similares a los obtenidos mediante el análisis en modo W. Además, esta mejora permitió aumentar la sensibilidad gracias a la inyección de mayores volúmenes.

En lo que respecta a la separación cromatográfica, se evaluó el uso de MeOH y ACN como disolventes orgánicos en la fase móvil obteniendo similares resoluciones cromatográficas pero con un ligero aumento de la sensibilidad cuando se eligió MeOH con HCOOH como modificador. Adicionalmente, se investigaron los efectos de la intensidad del lockmass, la relación de iones en la solución de calibración y el EDC en la exactitud de masa, consiguiendo las mejores condiciones posibles para el análisis instrumental.

La aportación más importante realizada en este trabajo es probablemente la posibilidad de realizar la identificación de los analitos encontrados en una misma inyección gracias a la adquisición en modo MS[E]. Este modo permite adquirir simultaneamente dos funciones: a baja y a alta energía de colisión. En la primera, suele favorecerse la ionización mediante la (des)protonación o la formación de aductos, facilitando la identificación del ion molecular y la asignación de una composición elemental o viceversa, es decir, obtener el/los iones predominantes a partir de la información estructural. En la segunda, se promueve la fragmentación de la molécula, lo cual es fundamental para la correcta identificación del analito. Entre diversas energías y rampas de energía de colisión, se seleccionó una rampa entre 15 y 40 eV como compromiso para los diferentes compuestos estudiados, aunque algunos de ellos presentaron, inevitablemente, una fragmentación pobre o nula. En cualquier caso, como se muestra en la figura 5 del artículo, se obtuvieron espectros muy similares a los obtenidos por MS/MS y, aunque menos selectivo, el modo MS[E] resulta ventajosos gracias a:

a) la obtención automática y simultánea de los espectros, sin necesidad de preseleccionar los compuestos de interés previamente a la inyección en el sistema,

permitiendo el análisis retrospectivo de las muestras sin necesidad de volver a analizarlas.

b) se conserva la distribución isotópica (*pattern isotópico*) de los fragmentos lo cual resulta útil en la identificación de los contaminantes y ayuda en su elucidación estructural

Como se muestra en la figura 6 del artículo, al comparar un positivo de carbendazima en un efluente urbano con los espectros incluidos en la librería doméstica creada, el grado de similitud (*match*) es muy alto, lo que implica un elevado grado de confianza en la correcta identificación del compuesto.

Este trabajo sienta las bases para el desarrollo de una metodología robusta y eficaz para el screening "universal" de contaminantes orgánicos mediante LC-HRMS en diferentes tipos de matrices. A partir de los resultados de este artículo, se establecieron las condiciones de análisis para futuras investigaciones.

Cabe remarcar la información disponible en la base de datos sobre los iones fragmentos obtenidos con masa exacta para más de 100 compuestos. Esto abre una vía para la creación de una librería de espectros a partir de datos obtenidos en diferentes laboratorios, la cual no esté centrada únicamente en una cierta familia de compuestos.

2.3. Artículo científico 2

"Target and non-target screening strategies for organic contaminants, residues and illicit substances in food , environmental and human biological samples by UHPLC-QTOF-MS"

R. Díaz, M. Ibáñez, J.V. Sancho and F. Hernández.

Analytical Methods, 4 (2012) 196-209.

Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF-MS†

Ramon Díaz, María Ibáñez, Juan V. Sancho and Félix Hernández*

In this paper, we illustrate the potential of ultra-high performance liquid chromatography (UHPLC) coupled with hybrid quadrupole time-of-flight mass spectrometry (QTOF MS) for large scale screening of organic contaminants in different types of samples. Thanks to the full-spectrum acquisition at satisfactory sensitivity, it is feasible to apply both (post)-target and non-target approaches for the rapid qualitative screening of organic pollutants in food, biological and environmental samples. Different strategies have been applied and compared in this work. The first approach consists of target screening based on automatically extracting the exact analyte masses with a narrow mass window (±10 mDa). The selection of analytes can be made after MS acquisition as non-specific analyte information is required when injecting the samples. The second, non-targeted approach, consists of a first component detection step followed by the search of the detected components in home-made spectral libraries. In this work, two types of libraries have been evaluated: a theoretical database, including the molecular formula of a large number of pollutants ($\sim$1000), and an empirical mass spectra library which includes a lower number of compounds for which reference standards were available. In all cases the confidence of the identification process was excellent, thanks to the value of information given in QTOF MS^E acquisition mode (*i.e.* simultaneous acquisition of low and high energy TOF MS spectra in a unique run). Both, target and non-target approaches, are complementary and both have advantages and drawbacks. Their application to different types of samples has allowed the detection of diverse organic compounds, for example the mycotoxin fumonisin B1 in food samples, cocaine and several metabolites in human urine, as well as several pesticides, antibiotics and drugs of abuse in urban wastewater.

Introduction

Nowadays, liquid chromatography (LC) hyphenated to mass spectrometry (MS) using a variety of mass analyzers is the technique of choice for the investigation of organic contaminants in most analyte/sample matrix combinations in environmental, food or toxicology fields. Mass analyzers used include triple quadrupole (QqQ),[1–6] time-of-flight (TOF), hybrid quadrupole time-of-flight (QTOF),[2,7–11] quadrupole-linear ion trap (QLIT)[3,12,13] or Orbitrap.[14] Many examples can be found in the literature dealing with pesticide residue analysis in environmental,[15] food[16,17] or biological samples,[18] using LC-MS based methods. Emerging contaminants, such as pharmaceuticals[1,5] or drugs of abuse,[19] amongst others, are increasingly being monitored in the environment by LC-MS because their medium-to-high polarity and low volatility make their determination fit better with LC. Similarly this applies to metabolites and transformation products, which are generally more polar than their parent molecules.

LC-tandem MS (LC-MS/MS) operating in Selected Reaction Monitoring mode (SRM) with QqQ analysers are the work-horses nowadays in target analysis.[20] LC-MS/MS methods rarely include more than two hundred analytes,[2,21,22] and with a few exceptions,[23] most of them are focused on a single family of contaminants. Excellent sensitivity and notable selectivity are achieved by LC tandem MS, allowing reliable quantification and identification of a considerable number of compounds. However, the presence of other contaminants that might be present in the samples would be ignored in LC-MS/MS under SRM mode (the most common approach), due to the analyte-specific information acquired. There is a need in the field of public health to develop reliable methods for large-scale screening that are capable of detecting and identifying a large number of hazardous compounds that can potentially be present in environmental and food samples. For this purpose, full spectrum acquisition techniques capable of providing accurate mass measurements are a great help.

Research Institute for Pesticide and Water, University Jaume I, Av. Sos Baynat S/N, 12071 Castellón, Spain. E-mail: felix.hernandez@qfa.uji.es; Fax: +34 964387368; Tel: +34 964387366
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c1ay05385j

Target and non-target screening strategies for organic contaminants, residues and illicit substances in food , environmental and human biological samples by UHPLC-QTOF-MS

R. Díaz, M. Ibáñez, J.V. Sancho and F. Hernández

Research Institute for Pesticides and Water, Universitat Jaume I, E-12071 Castelló, Spain

ABSTRACT

In this paper, we illustrate the potential of ultra-high performance liquid chromatography (UHPLC) coupled with hybrid quadrupole time-of-flight mass spectrometry (QTOF MS) for large scale screening of organic contaminants in different types of samples. Thanks to the full-spectrum acquisition at satisfactory sensitivity, it is feasible to apply both (post)-target and non-target approaches for the rapid qualitative screening of organic pollutants in food , biological and environmental samples. Different strategies have been applied and compared in this work. The first approach consists of target screening based on automatically extracting the exact analyte masses with a narrow mass window (±10 mDa). The selection of analytes can be made after MS acquisition as non-specific analyte information is required when injecting the samples. The second, non-targeted approach, consists of a first component detection step followed by the search of the detected components in home-made spectral libraries. In this work, two types of libraries have been evaluated: a theoretical database, including the molecular formula of a large number of pollutants ($\sim$1000), and an empirical mass spectra library which includes a lower number of compounds for which reference standards were available. In all cases the confidence of the identification process was excellent, thanks to the value of information given in QTOF MS^E acquisition mode (*i.e.* simultaneous acquisition of low and high energy TOF MS spectra in a unique run). Both, target and non-target approaches, are complementary and both have advantages and drawbacks. Their application to different types of samples has allowed the detection of diverse organic compounds, for example the mycotoxin fumonisin B1 in food samples, cocaine and several metabolites in human urine, as well as several pesticides , antibiotics and drugs of abuse in urban wastewater.

INTRODUCTION

Nowadays, liquid chromatography (LC) hyphenated to mass spectrometry (MS) using a variety of mass analyzers is the technique of choice for the investigation of organic contaminants in most analyte /sample matrix combinations in environmental, food or toxicology fields. Mass analyzers used include triple quadrupole (QqQ),[1–6] time-of-flight (TOF), hybrid quadrupole time-of-flight (QTOF),[2,7–11] quadrupole-linear ion trap (QLIT)[3,12,13] or Orbitrap .[14] Many examples can be found in the literature dealing with pesticide residue analysis in environmental,[15] food [16,17] or biological samples,[18] using LC-MS based methods. Emerging contaminants, such as pharmaceuticals [1,5] or drugs of abuse,[19] amongst others, are increasingly being monitored in the environment by LC-MS because their medium-to-high polarity and low volatility make their determination fit better with LC . Similarly this applies to metabolites and transformation products, which are generally more polar than their parent molecules.

LC-tandem MS (LC-MS/MS) operating in Selected Reaction Monitoring mode (SRM) with QqQ analysers are the workhorses nowadays in target analysis.[20] LC-MS/MS methods rarely include more than two hundred analytes ,[2,21,22] and with a few exceptions,[23] most of them are focused on a single family of contaminants. Excellent sensitivity and notable selectivity are achieved by LC tandem MS , allowing reliable quantification and identification of a considerable number of compounds. However, the presence of other contaminants that might be present in the samples would be ignored in LC-MS/MS under SRM mode (the most common approach), due to the analyte -specific information acquired. There is a need in the field of public health to develop reliable methods for large-scale screening that are capable of detecting and identifying a large number of hazardous compounds that can potentially be present in environmental and food samples. For this purpose, full spectrum acquisition techniques capable of providing accurate mass measurements are a great help.

To solve the limitations of unit resolution mass spectrometers , two main alternatives, based on the use of high-resolution MS instruments, are of note at present: the time-of-flight [24] and Orbitrap [25,26] analysers. Both provide full spectrum accurate-mass data at satisfactory sensitivity. These capabilities are very helpful for detecting and identifying not only priority known pollutants but many other unknown contaminants that might be a risk for human health.[27–29]

Although quantitative applications have been reported using LC-TOF MS or LC-QTOF MS ,[8,10,30] quantification does not seem to be the most attractive feature of these analysers. This may be due to the higher limits of detection and narrower linear dynamic range in comparison to QqQ analysers. One of the most interesting applications of TOF MS deals with the wide-scope screening of a large number of contaminants and residues in different types of

samples, as that allows a significant amount of useful information on ionisable compounds present in the sample to be obtained.[31] Generic (universal) sample treatments and chromatographic separations are required to broaden the scope of the method to as many compounds as possible. Besides, the elevated acquisition speed of TOF makes it compatible with ultra-high pressure (Ultra-Performance) liquid chromatography (UHPLC/UPLC). This technique provides fast, high-resolution separations that will hopefully minimize matrix effects and render high mass spectra purity, improving the screening process.

Different strategies can be used to extract analytical information from full-acquisition accurate mass data. A genuine non-target analysis involves the automated component detection from the total ion chromatogram (TIC) and the mass spectra deconvolution for a subsequent comparison with mass spectral libraries. Nevertheless, electrospray ionization (ESI) is not an ion source as stable and reproducible as electron ionization,[32] and commercial, standardized ESI mass spectra library are not available. Instead, theoretical mass spectra libraries, based on the molecular formula database, can be built which facilitate increasing the number of compounds that can be searched. These use accurate mass measurements and isotopic pattern information for identification. Home-made empirical libraries can also be used, but these normally include much fewer compounds due to the need to inject standards. These experimental libraries offer fragmentation and retention time information as well, providing more confidence in the compound identification process.[33] However, the possibility of detecting and identifying the sample contaminants, using both mass spectra libraries in a non-target analysis, depends on the success of the deconvolution process, *i.e.* the capability of the software to find the component peaks and to obtain mass spectra as free as possible of sample interferents. Obviously, the more complex the matrix, the more difficult the deconvolution will be.

An efficient approach to overcome the component detection limitations is the use of "post-target" methodology,[34,35]*i.e.* the selection of the analytes to be searched is done after MS acquisition. A post-target screening facilitates the detection of the compounds as it is only focussed on those pollutants selected. It is unnecessary to totally deconvolute all components present in the samples, these mainly belong to matrix compounds. Furthermore, processing and reviewing steps become easier as fewer compounds are searched for and consequently detected.

In this work three sample types have been selected (wastewater, food and human urine) to explore the potential of UHPLC-(Q)TOF MS to detect and identify/elucidate organic contaminants and/or residues. Two strategies have been applied for this purpose: a post-target screening, based on mass filtering at the exact mass of the compound investigated (typically the (de)protonated molecule) using narrow mass extractionwindows and a non-target methodology

using both empirical and theoretical mass spectra libraries. QTOF MS has been used under MS^E mode, *i.e.* simultaneous acquisition at low (LE) and high collision energy (HE) functions, which provides useful information on the (de)protonated molecules (commonly at LE) and on the main fragments ions (commonly in HE). On the basis of this information, and on isotopic distribution observed in the spectra , the reliable identification of the compounds detected in the samples was feasible.

EXPERIMENTAL

Reagents and chemicals

HPLC-grade water was obtained from deionized water passed through a Milli-Q water purification system (Millipore, Bedford, MA, USA). HPLC-grade methanol (MeOH) and acetonitrile (ACN) were purchased from ScharLab (Barcelona, Spain). Formic acid (HCOOH) (>98%) was obtained from Fluka (Buchs, Switzerland). Sodium hydroxide (>99%) was obtained from ScharLab. Leucine enkephalin, used as lock mass, was purchased from Sigma Aldrich (St Louis, MO, USA).

Reference compounds were purchased from Acros Organics (Geel, Belgium), Bayer Hispania (Barcelona, Spain), Fort Dodge Veterinaria (Gerona, Spain), Vetoquinol Industrial (Madrid, Spain), Aventis Pharma (Madrid, Spain), Sigma Aldrich (St Louis, MO, USA), Cerilliant (Round Rock, TX, USA), Dr Ehrenstorfer (Augsburg, Germany), Riedel-de Haën (Seelze, Germany), the National Measurement Institute (Pymble, Australia) and Fluka. All reference materials presented purity higher than 93% (w/w).

Instrumentation

An UPLC Acquity system coupled with a hybrid quadrupole orthogonal acceleration-time-of-flight (Q-oaTOF) mass spectrometer (QTOF Premier, Waters, Milford, MA) provided with an orthogonal Z-spray lockspray electrospray interface (ESI) was used.

Mobile phases A and B were water and methanol respectively, both with 0.01% formic acid. The separation was performed on an Acquity C18 BEH analytical column (150 mm × 2.1 mm, i.d. 1.7 μm) at a flow rate of 0.3 mL min^{-1} (at 60 °C). The initial percentage of methanol was 10%, which was linearly increased to 90% in 14 min, followed by a 2 min isocratic period and, then, returned to initial conditions during 2 min in total run duration of 18 min. The injection volume was 50 μL.

Cone and nebulizer gas were nitrogen (Praxair, Valencia, Spain) at flow rates of 60 L h⁻¹ and 600 L h⁻¹, respectively. The nitrogen desolvation temperature was set to 350 °C and the source temperature to 120 °C. A cone voltage of 25 V and capillary voltages of 3.5 kV and 2.5 kV in positive and negative ionisation modes, respectively, were used.

TOF MS resolution was ~10 000 at full width half maximum (FWHM) in V-mode. MS spectra were acquired over an m/z range 50–1000. Collision gas was argon 99.995% (Praxair, Valencia, Spain), which was always turned on with a pressure of approximately 5×10^{-3} mbar.

Two acquisition functions were created with different collision energies. The first one, the low energy (LE) function, at low collision energy (4 eV) and the second one, the high energy (HE) function, with a collision energy ramp ranging from 15 to 40 eV. The scan time values of LE and HE functions were set to 0.2 and 0.15 s, respectively, both with an inter-scan delay of 0.05 s.

The lock mass (leucine enkephalin, 2 mg L⁻¹ in ACN : water, 50 : 50) was introduced *via* the lock spray needle at a flow rate of 30 μL min⁻¹ using a reagent manager pump (Waters). A cone voltage of 60–70 V was selected and checked daily to obtain adequate signal intensity for this compound (around 500 counts).

Calibration of the m/z-axis was performed using the built-in single-syringe pump , directly connected to the interface. Calibration from 50 to 1000 m/z was conducted with a 1 : 1 mixture of 0.05 M NaOH : 5% HCOOH diluted (1 : 25) with water/ACN (20 : 80 v/v) plus imazalil (m/z 297.0561) at a final concentration of 500 μg L⁻¹.

Data station operating software was MassLynx v 4.1. ChromaLynx XS application manager was used for non-target (deconvolution and library search) as well as for target analysis.

Sample treatment

8 wastewater samples—4 influent (IWW) and 4 effluent (EWW)—10 human urine and 6 food samples (2 oranges, 2 banana and 2 corn samples) were analysed for comparing the screening approaches.

50 mL of wastewater were pre-concentrated by off-line SPE using 200 mg Oasis HLB cartridges, eluted with 5 mL of MeOH, evaporated under a gentle nitrogen stream at 40 °C and reconstructed with 1 mL water : MeOH (90 : 10 v/v).

Food sample extraction was performed according to previous work developed by our group .[6,36] 20 g of triturated and homogenized orange or banana samples were extracted with 60 mL water : MeOH (20 : 80 v/v) for 2 min using a high-speed blender, filtered and diluted

with water : MeOH (20 : 80 v/v) to a final volume of 100 mL. Afterwards, an aliquot of the extract was diluted eightfold with water.

2.5 g of crushed corn sample were extracted with 10 mL ACN : water (80 : 20 v/v) with 0.1% HCOOH and mechanically shaken for 90 min.[6] Afterwards, the solution was centrifuged, and a 5 mL aliquot of supernatant was diluted twofold with water.

Human urine samples from healthy volunteers and from people involved in drug detoxification programmes were centrifuged, diluted fivefold with water and directly injected into the LC-QTOF instrument.

Software parameters

The deconvolution and spectra rejection parameters were selected as follows:

- minimum peak width at 5% height: 4 s,
- peak-to-peak baseline noise: 5,
- smoothing activated,
- mass tolerance (mass window width): 20 mDa,
- two mass chromatograms extracted for each component in LE function (5 mass chromatograms for HE), *i.e.* 2 or 5 coeluting ions to be extracted with the narrow window mass selected (±10 mDa).

The values for minimum peak width and mass window were selected as a function of the chromatographic resolution and mass accuracy data of our instrument.

Accurate mass scoring parameters were selected as follows:

- Number of ions used for accurate mass scoring: 2
- Minimum intensity (% of largest peak in the range): 10
- High precision mass tolerance (colouring in green): 2.5 mDa
- Low precision mass tolerance (colouring in yellow) = 5 mDa

These values were selected according to our own experience and characteristics of the LC-QTOF MS equipment used, but they might be modified according to the performance of the instrument used in each laboratory.

RESULTS AND DISCUSSION

Mass resolving power is an important issue for the correct detection and identification of the suspect compounds. Even if the 10 000 at 10% valley resolution (20 000 FWHM) required by the EC Decision 2002[37] is not achieved by the (Q)TOF mass spectrometer used, we consider that mass accuracy is really the key in the identification of the compounds in the wide-scope screening. Although strongly correlated, mass resolving power and mass accuracy are not strictly the same. In a previous work,[33] improving the resolution (about 18 000 FWHM) by doubling the path length using the so-called W-mode in different matrices (influent and effluent wastewater, surface water, pepper and cucumber) showed no significant effect for the compounds tested on mass accuracy achieved using UHPLC separation. On the other hand, a 20 mDa mass window has been used in this work for both, non-target and post-target strategies, as a compromise between ensuring correct chromatographic peak at both ends and attainable selectivity. Lower mass windows (*e.g.* 5 and 10 mDa) were also tested, but finally discarded as no satisfactory chromatographic peaks were always ensured. Newer instruments with stable mass accuracy across the peak could facilitate the screening process by reducing this mass window, even down to 1 mDa.

The non-target and post-target strategies studied in this work were applied to all selected samples to test the screening capabilities and for comparison purposes. A flowchart of the process is shown in Fig. 1.
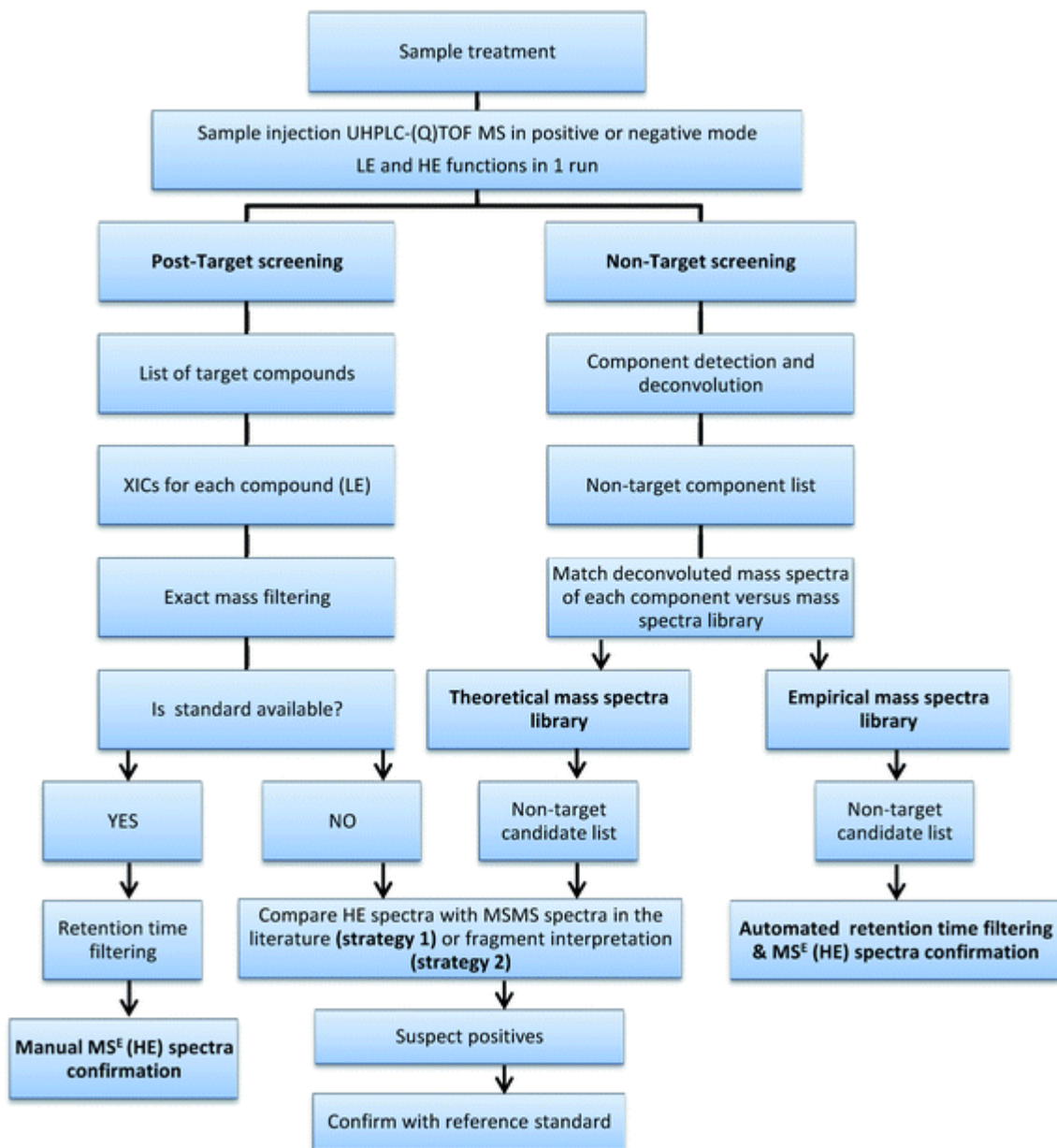
Fig. 1 Flowchart of the overall screening process.

1. Non-target screening

A true non-target screening using LC-(Q)TOF MS is a challenging task as it is very difficult to detect and identify trace level contaminants when no selection is made on the compounds to

be searched.[38] In this work, non-target screening was applied to environmental, food and biological samples to evaluate the potential of the algorithm to detect components when dealing with complex matrices. For this purpose, the deconvolution software ChromaLynx XS in a non-target mode was used. The software applies a component detection algorithm (CODA) to deconvolute the TIC and detect the components present in the sample. Afterwards, it compares the spectra assigned to every component with those included in the home-made libraries. To facilitate the confirmation of the identity of the components detected, two functions were simultaneously acquired at different collision energies (MS^E). The LE function was used to obtain the (de)protonated molecules (occasionally adducts and fragment ions). The HE function was used to promote fragmentation, improving the identification of the positive findings as spectra obtained were quite similar to those of MS/MS experiments.[33,39] This acquisition provides reproducible spectra without the need of precursor ion pre-selection in the first quadrupole. The success for detecting and identifying non-target compounds using this approach obviously depends on the deconvolution process. In addition, MS^E provides not only fragmentation spectra but also isotopic pattern information of the fragments and it conserves adduct and/or dimer information. However, two main limitations were noticed when MS^E was applied to non-target screening:

(a) As there is no pre-selection of precursor ion in the quadrupole, the MS^E approach is less specific and might be conflictive when dealing with non-selective fragments in the presence of co-eluting related compounds. This occurs, for example, when investigating amphetamine -like compounds amphetamine and methamphetamine. As can be seen in Fig. 2, both drugs elute at very close retention times and present poor and identical HE spectra , with the most abundant ion being the non-selective fragment at *m/z* 91 corresponding to tropylium ion. Moreover, as protonated molecules have relatively poor abundance in the LE function (especially amphetamine) it could be very difficult to distinguish both compounds at low concentration levels.
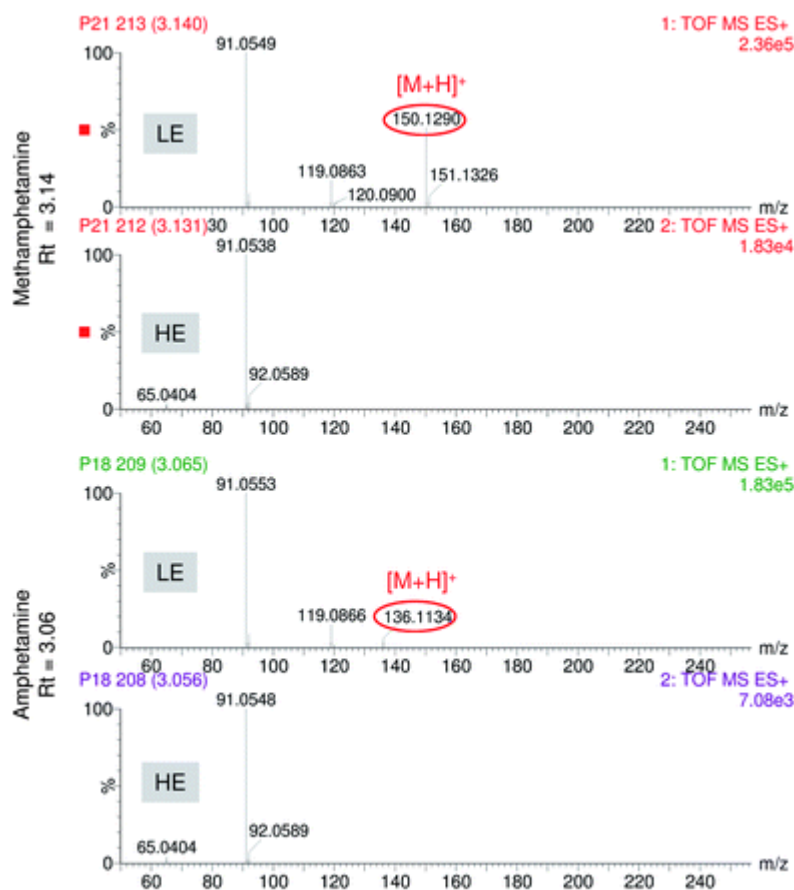
Fig. 2 LE and HE mass spectra for amphetamine-like compounds. LE spectra for methamphetamine and amphetamine show notable in-source fragmentation with different [M + H]$^+$ ion (*m/z* 150.1290 and 136.1134). Identical HE spectra (*m/z* 91.0548 corresponding to tropylium ion) are obtained for both compounds.

(b) The success of the MS^E approach can be limited by the quality of the spectrum.[39] Thus, low sensitivity or strongly interfered spectra end up making it unfeasible to match with library spectra as well as not being able to elucidate the component using fragment interpretation. In these cases, additional MS/MS experiments would be helpful in the identification/elucidation process.

Finally, the software returns a match factor for the comparison of standard and candidate mass spectra and gives the mass errors for the 2 most abundant ions present in the LE function and for the 5 main fragment ions present in the HE function. A positive match can be

filtered by a minimum match factor and retention time, if available. In this work, a relatively low match factor (in reverse fit) of 70% was selected as a compromise. This facilitated the reviewing process of positives without losing potential hazardous compounds that could be present in the samples although with low match factors. Two types of mass spectra libraries were evaluated in this work as discussed in the following sections.

When no or unsatisfactory match is obtained, the components appear to be tentative. In these cases, the elucidation of the compound requires a lot of time and effort with a low possibility of success. Furthermore, the majority of non-matched components are likely to be matrix compounds.

1.1 Theoretical library.

Initially, a database containing approximately one thousand pollutants of different families   (pesticides , antibiotics , pharmaceuticals , illicit drugs , mycotoxins , anabolic steroids , personal care products and metabolites) was built (see ESI†). The compounds were included based on our own experience on LC-MS/MS analysis of environmental and food samples, and on bibliographic data on LC-MSamenable organic pollutants. The database was created separating positive and negative ionisable compounds. It contained information on the molecular formula (required by the software), exact mass of the neutral and the (de)protonated molecule, as well as supplementary information† of the compound type and on retention time, when available. From the molecular formulae of each compound, two theoretical mass spectra libraries (for positive and negative ionisation modes) were automatically built, containing theoretical nominal mass spectra of the (de)protonated molecule and sodium adducts as well as the theoretical isotopic pattern expected for each compound. Each library (positive and negative modes) was used in the corresponding acquisition mode.

A drawback of the theoretical library (and also of the empirical mass spectra library) is that TOF MS spectra are stored in nominal mass for NIST format compatibility, and in this step the mass accuracy information given by TOF MS is lost. In order to minimize this limitation, the mass errors between the measured masses of the compound detected and the exact masses of the candidates formulae are calculated and used in a subsequent step, to rank them and to propose the most plausible identity (accurate mass scoring).[35]

1.2 Empirical library.

Details about reference standards injected and conditions for the creation of the empirical spectra library are reported in Díaz *et al.*[33] Briefly, around 230 reference standards of selected contaminants were injected in both, positive and negative, ionisation modes at low and high collision energy (MS^E mode) to obtain retention time and fragmentation information under the previously optimized conditions.[33] For each compound, two library entries (LE and HE spectra) were created including name, exact mass, retention time and spectra .

Detection/identification problems derived from LE adducts formation and/or important in-source fragmentation were prevented by analyzing the samples under exactly the same conditions as the reference standards. This favoured the task and minimized the risk of potential false negatives. Furthermore, HE mass spectra were automatically matched with those included in the empirical library which greatly facilitated the confirmation of the compound identity. In our experience, HE provided highly reproducible spectra (independently of the type of sample analysed) when the component was found at relatively high abundance. As signal intensity is the main limitation during the component detection step, HE spectra facilitated identification of the compound when its spectrum was available in the library in those components detected by the non-target approach.

2. Post-target screening

Trying to avoid the dependence of the screening success on the component detection algorithm, a post-target screening strategy was applied including an extraordinarily large number of compounds in the search. The term "post-target" was first used by our group [34,35,40] as a target screening without pre-selection of the analytes before analysis. It consists of searching for a list of target compounds after MS full-acquisition. Other authors name this approach, when reference standards are unavailable, as suspect screening.[20] In the post-target screening, a database with the same compounds included in the theoretical library of the non-target approach was used (ESI†). ChromaLynx XS uses the molecular formula to calculate the exact mass for $[M + H]^+$. Then, the software automatically performs the extraction of a nw-XIC (20 mDa) for each compound in the LE and HE functions and looks for peaks (*S/N* and peak width higher than pre-selected values) in the corresponding chromatogram . A list of potential candidates found in the sample is shown in different colours depending on accurate mass measurement; positive (green) for error <2.5 mDa, tentative (yellow) for error between 2.5 and 10 mDa, and negative (red) for error >10 mDa. Furthermore, as in the non-target approach, ChromLynx XS filters positive findings according to retention time deviation limit when this information is available in the database (reference standards previously injected). The retention time window was set in ±0.5 min but

accepted tolerance was 2.5%. Thus, the retention times for 231 analytes , injected when building the empirical library, were also introduced in the database. In this way, nw-XIC, top peak spectra and mass error as well as isotopic distribution fit (i-FIT) information, retention time (measured and expected when already known) and peak area were available for positive matches. Fig. 3 shows the ChromaLynx XS browser for a positive of mycotoxin Fumonisin B1 in a corn sample using this approach.
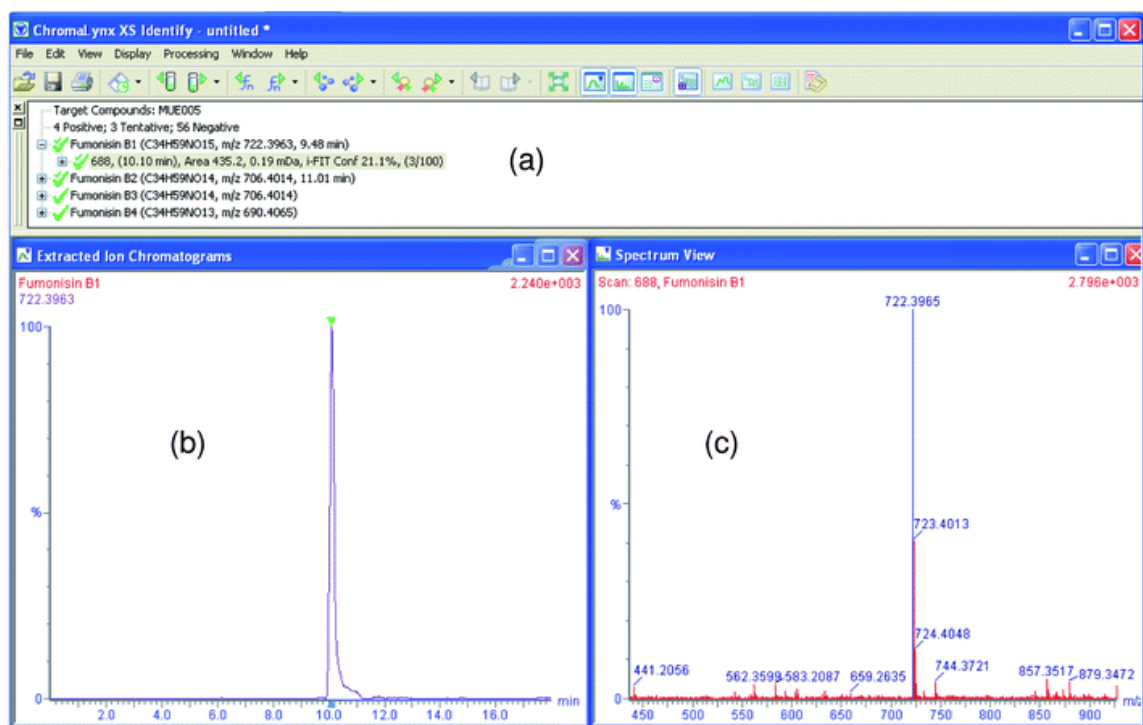
Fig. 3 ChromaLynx XS browser with accurate mass confirmation for Fumonisin B1 in corn using post-target screening. (a) Candidate list for compound with mass error <2.5 mDa (which offers retention time, area, mass error and i-FIT information), (b) nw-XIC for suspected candidate (at 20 mDa window). (c)mass spectrum (in blue, candidate peak is shown).

QTOF MS post-target screening has proved to be an efficient tool due to the high number of pollutants screened. The potential of this approach to detect different families of organic contaminants, for example drugsof abuse or antibiotics in environmental samples, has been reported recently.[39,41] The easy reviewing step and the relevant information obtained, such

as accurate mass spectrum of the peak, mass error for the protonated molecule and the most abundant fragments, and isotopic distribution, give high confidence to the confirmation of potential positives even without reference standards being available. The large number of contaminants included in the list (more than 1000) opens a new scenario in screening, favouring a more realistic overview when investigating organic contaminants in different applied fields. However, if only the predicted presence of the protonated molecule was taken into account in the LE function, potential in-source fragments would not be detected (*e.g.* as occurs in amphetamine -like compounds, see Fig. 2), not even sodium or other adducts that could be formed.

In this work, formic acid was added to the mobile phases . Under this situation, ammonia adducts and other adducts like $[M + MeOH + H]^+$ or $[M + K]^+$ would not normally be expected. However, sodium adducts are common for many LC-amenable compounds, and they might be present despite using formic acid. In our own experience, 88 out of 231 compounds (38%) included in the experimental library showed sodium adducts at relative abundance higher than 10%. Among them, 38 compounds (16% of the total compounds) presented the $[M + H]^+$ ion at relative intensity lower than 10%, this becoming the $[M + Na]^+$, the most abundant ion in the mass spectra . When analyzing real samples, sodium adducts might be found at higher abundance due to the normal presence of sodium in the sample matrices. Therefore, it is important to include sodium adducts in the screening to avoid potential false negatives in those cases where it is the most abundant ion (see Table 1). However, it seems reasonable not to include sodium adducts for all analytes investigated, as the processing and the reviewing step would be much longer and more tedious. The injection of reference standards and/or literature search, along with analyst experience, are necessary parts of knowing when it is reasonable to include compound adducts to improve the confidence of the screening process.

A similar problem may occur when important in-source fragmentation takes place at the LE function. In this work, we used 25 V cone voltage as better sensitivity was observed for selected analytes in the 20–30 V range.[33] Obviously, this cone voltage is a compromise value as it is not the best choice for all compounds but it is impossible to optimize any variable for all LC-amenable compounds included in the database. As previously stated, amphetamine ($[M + H]^+$ 136.1126) presents an in-source fragment at *m/z* 91.0553 as the most abundant ion in the spectra , while the protonated molecule has an abundance lower than 10% (Fig. 2). Other examples are insecticide carbaryl (fragment at *m/z* 145.0563) or pesticide metabolite aldicarb sulfoxide (fragment at *m/z* 89.0415). In these cases, analyte detection in samples based on testing $[M + H]^+$ presence would be only feasible at relatively high analyte concentrations.

Table 1 Results of different screening approaches for representative samples. Common data for all approaches are shown: retention time (RT), retention time deviation (only when available) and mass error. For target screening, confirmation with MS^E was performed when reference standard was previously injected. Detection of sodium adduct and/or fragment ions in the LE function is reported. Reverse Fit is given for compounds detected by non-target approaches

Effluent wastewater				Target Screening					Non-target screening	
				Tentative Identification		Confirmed with standard?	LE function		Theoretical library	Empirical library
Compound	RT (min)	ΔRT (%)	Δmass (mDa)	Strategy 1[a]	Strategy 2[b]		Na adduct?	Fragment ions?	Match reverse fit	Match reverse fit
Antipyrine	5.43	-	1.4			-				
Bamethan	3.43	-	0.5	✓	×	-				
Bisoprolol	7.25	-	0.0	✓	✓	-			929	
Caffeine	4.19	-	0.2	✓	✓	-				
Carbendazim	4.45	0.04	0.5			yes		✓(1)	923	898
Celiprolol	6.41	-	0.7	✓	✓	-				
Clarithromycin	10.11	0.10	1.7			yes			855	
Clofibric acid	12.60	-	2.5	✓	×	-				
Codeine	2.82	0.00	1.6			yes			943	794
Diazinon	13.04	0.01	0.9			yes				
Diuron	9.98	0.02	0.3			yes				
Erithromycin(-H2O)	9.86	0.01	1.2			yes				
Gabapentin	3.47	0.09	0.5	✓	✓	yes		✓(2)	903	
Irbesartan	11.42	0.10	0.0	✓	✓	yes			953	
Ketoprofen	10.61	0.04	0.5			yes	✓			
Metoprolol	5.55	-	0.0	✓	✓	-				
Nordiazepam	11.01	-	0.1		✓	-				
OD-PABA	5.49	-	0.9	✓	✓	-			953	
Oxazepam	10.19	-	0.4		✓	-				
Oxprenolol	2.77	-	0.2	✓	✓	-				
Propylphenazone	9.29	-	0.4	✓	✓	-				
Terbutryn	11.69	0.14	1.0			yes				

Thiabendazol	5.22	0.12	0.1			yes				
Trimethoprim	3.83	0.10	0.9			yes			980	925
Valsartan	11.59	0.05	1.9	✓	✓	yes	✓		849	
Venlafaxine	7.17	0.46	0.3			yes		✓(1)		
MDMA	3.78	0.00	0.0			yes		✓(2)		
Bezafibrate*	11.07	0.01	2.1			yes	✓			
Gemfibrozil*	13.84	0.03	2.5			yes	✓			

*Compounds found only as sodium adduct ion

Orange				Target Screening					Non-target screening	
				Tentative Identification		Confirmed with standard?	LE function		Theoretical library	Empirical library
Compound	RT (min)	ΔRT (%)	Δmass (mDa)	Strategy 1[a]	Strategy 2[b]		Na adduct?	Fragment ions?	Match reverse fit	Match reverse fit
Imazalil	9.22	0.03	0.3			yes			916	925
Thiabendazol	5.21	0.03	0.3			yes			893	916

Banana peel				Target Screening					Non-target screening	
				Tentative Identification		Confirmed with standard?	LE function		Theoretical library	Empirical library
Compound	RT (min)	ΔRT (%)	Δmass (mDa)	Strategy 1[a]	Strategy 2[b]		Na adduct?	Fragment ions?	Match reverse fit	Match reverse fit
Chlorpyrifos	14.61	0.01	0.5			yes	✓			
Diazinon	13.03	0.01	0.7			yes			854	863
Imazalil	9.21	0.02	0.9			yes			931	923

Corn				Target Screening					Non-target screening	
				Tentative Identification		Confirmed with standard ?	LE function		Theoretical library	Empirical library
Compound	RT (min)	ΔRT (%)	Δmass (mDa)	Strategy 1[a]	Strategy 2[b]		Na adduct?	Fragment ions?	Match reverse fit	Match reverse fit
Fumonisin B1	10.10	0.03	0.8			yes			785	800
Fumonisin B2	11.69	0.03	1.2			yes			681	686
Fumonisin B3	10.97	-	1.3	✓	✓	-				
Fumonisin B4	12.44	-	0.2	✓	✓	-				

Urine sample				Target Screening					Non-target screening	
				Tentative Identification		Confirmed with standard ?	LE function		Theoretical library	Empirical library
Compound	RT (min)	ΔRT (%)	Δmass (mDa)	Strategy 1[a]	Strategy 2[b]		Na adduct?	Fragment ions?	Match reverse fit	Match reverse fit
Gabapentin	3.27	0.05	0.8			yes		(2)	893	
Nicotine	2.72	-	1.1	✓	✓	-				
Paracetamol	2.68	0.02	0.3			yes				
Risperidone	6.72	0.10	0.0			yes				
Amphetamine	3.27	0.20	0.4			yes		(2)	874	
Benzoylecgonine	5.04	0.02	1.0			yes			895	967
Cocaethylene	6.25	0.11	0.4			yes				
Cocaine	5.16	0.19	0.8			yes			716	
norBenzoylecgonine	5.28	0.00	0.6			yes				
norCocaethylene	5.71	-	0.7	✓	✓	-				
norCocaine	5.34	0.05	0.2	✓	✓	-				

[a] *Strategy 1, used for tentative identification of the compounds when the standard was not available, consisted of comparing the main fragments observed in the HE function with common MS/MS product ions reported in the literature.*

[b] *Strategy 2 was made by justifying the HE accurate mass fragments using a bond-disconnecting software*

Other compounds, like anabolic steroids , are frequently ionised forming adducts with MeOH, acetonitrile, ammonium or sodium (as a function of the mobile phase and sample matrix composition) and/or they suffer in-source fragmentation with neutral losses of one, or even two, water molecules ($[M − H_2O + H]^+$, $[M − 2H_2O + H]^+$).[42] The later drawback is more difficult to solve than adducts formation, but it could be circumvented by including empirical formula of the known fragment ions in the database. Again, information reported in the literature and/or from reference standards injection would be required to include expected fragments in the database. Although fragmentation behaviour is not completely known in most cases, in our experience, this effect is less common than adduct formation. Indeed, only 6 out of 231 compounds (3% of the compounds studied) almost exclusively presented the fragment ion as base peak, with the protonated ion being practically absent. In these particular cases, monitoring this fragment is mandatory for compound detection. In-source fragmentation turns into a useful confirmatory tool when the reference standard is available and/or its behaviour is well known. Thus, including most abundant fragments is always useful for automated confirmation.

As an illustrative example, Table 2 shows information on database entries for different types of analytes included in this work. The molecular formula of the ion, when adduct formation and/or in-source fragmentation occurred, was also introduced in the database, as well as the bibliographic source, when information on possible occurrence of these ions was not directly obtained from reference standard injection.

Table 2 Illustrative example of the database used in the post-target approach. Database entries were created including molecular formula, retention time (when available), accurate mass and pollutant family, as well as bibliographic source when data were not empirically obtained by reference standard injection. Different entries were created for sodium adducts (marked as Na) and in-source fragments (marked as F1, F2, *etc.*). Entries marked with (*) indicate the main ion/s in the reference standard spectrum.

Compound	Molecular formula	RT (min)	Ion type	Accurate mass	Pollutant family	Source
Oxytetracycline*	$C_{22}H_{24}N_2O_9$	4.83	$[M + H]^+$	461.1560	Antibiotic	
Oxytetracycline F1	$C_{22}H_{22}N_2O_8$	4.83	Fragment ion	443.1454	Antibiotic	
Oxytetracycline F2	$C_{22}H_{19}NO_8$	4.83	Fragment ion	426.1189	Antibiotic	
Amphetamine	$C_9H_{13}N$	3.06	$[M + H]^+$	136.1126	Illicit drug	
Amphetamine F1	C_9H_{10}	3.06	Fragment ion	119.0861	Illicit drug	
Amphetamine F2*	C_7H_6	3.06	Fragment ion	91.0548	Illicit drug	
MDMA	$C_{11}H_{15}NO_2$	3.14	$[M + H]^+$	194.1181	Illicit drug	
MDMA F1*	$C_{10}H_{10}O_2$	3.14	Fragment ion	163.0759	Illicit drug	
MDMA F2	$C_8H_6O_2$	3.14	Fragment ion	135.0446	Illicit drug	
6-OH-4-Cl-dehydromethyltestosterone	$C_{20}H_{27}O_3Cl$	9.91	$[M + H]^+$	351.1727	Steroid	
6-OH-4-Cl-dehydromethyltestosterone ($-H_2O$)	$C_{20}H_{25}O_2Cl$	9.91	Fragment ion	333.1621	Steroid	
6-OH-4-Cl-dehydromethyltestosterone ($-2 \times H_2O$)	$C_{20}H_{23}OCl$	9.91	Fragment ion	315.1515	Steroid	
6-OH-4-Cl-dehydromethyltestosterone (Na)*	$C_{20}H_{26}NaO_3Cl$	9.91	$[M + Na]^+$	373.1547	Steroid	
Ethisterone	$C_{21}H_{28}O_2$	—	$[M + H]^+$	313.2168	Steroid	JMS,42,2007,497-516
Ethisterone [M+Na+MeOH]+	$C_{22}H_{31}NaO_3$	—	$[M + Na + MeOH]^+$	367.2249	Steroid	JMS,42,2007,497-516

Compound	Molecular formula	RT (min)	Ion type	Accurate mass	Pollutant family	Source
Fumonisin B1*	$C_{34}H_{59}NO_{15}$	10.13	$[M + H]^+$	722.3963	Mycotoxin	
Fumonisin B2*	$C_{34}H_{59}NO_{14}$	11.66	$[M + H]^+$	706.4014	Mycotoxin	
Aldicarb sulfoxide	$C_7H_{14}N_2O_3S$	3.19	$[M + H]^+$	207.0803	Pesticide	
Aldicarb sulfoxide (Na)*	$C_7H_{13}NaN_2O_3S$	3.19	$[M + Na]^+$	229.0623	Pesticide	
Aldicarb sulfoxide F1*	C_4H_8S	3.19	Fragment ion	89.0351	Pesticide	
Aldicarb sulfoxide F2	C_5H_9NOS	3.19	Fragment ion	132.0483	Pesticide	
Tebufenozide	$C_{22}H_{28}N_2O_2$	12.54	$[M + H]^+$	353.2229	Pesticide	
Tebufenozide (Na)*	$C_{22}H_{27}N_2O_2Na$	12.54	$[M + Na]^+$	375.2048	Pesticide	
Tebufenozide (2M + Na)*	$C_{44}H_{55}N_4O_4Na$	12.54	$[2M + Na]^+$	727.4199	Pesticide	
Tebufenozide F1*	$C_{18}H_{20}N_2O_2$	12.54	Fragment ion	297.1603	Pesticide	
Tebufenozide F2*	C_9H_8O	12.54	Fragment ion	133.0653	Pesticide	
Azinphos-methyl	$C_{10}H_{12}N_3O_3PS_2$	10.49	$[M + H]^+$	318.0136	Pesticide	
Azinphos-methyl (Na)*	$C_{10}H_{11}N_3O_3PS_2Na$	10.49	$[M + Na]^+$	339.9956	Pesticide	
Azinphos-methyl F1	$C_8H_5N_3O$	10.49	Fragment ion	160.0511	Pesticide	
Azinphos-methyl F2*	C_8H_5NO	10.49	Fragment ion	132.0449	Pesticide	
Azoxystrobin	$C_{22}H_{17}N_3O_5$	10.97	$[M + H]^+$	404.1246	Pesticide	
Azoxystrobin (Na)*	$C_{22}H_{16}NaN_3O_5$	10.97	$[M + Na]^+$	426.1066	Pesticide	
Azoxystrobin F1*	$C_{21}H_{13}N_3O_4$	10.97	Fragment ion	372.0984	Pesticide	
Bifenazate	$C_{17}H_{20}N_2O_3$	11.92	$[M + H]^+$	1.0078	Pesticide	
Bifenazate (Na)*	$C_{17}H_{19}N_2O_3Na$	11.92	$[M + Na]^+$	22.9898	Pesticide	
Bifenazate F1	$C_{13}H_{11}NO$	11.92	Fragment ion	198.0919	Pesticide	
Bifenazate F2	$C_{12}H_{11}N$	11.92	Fragment ion	170.0970	Pesticide	
Dimethoate	$C_5H_{12}NO_3PS_2$	5.76	$[M + H]^+$	230.0075	Pesticide	
Dimethoate (Na)*	$C_5H_{11}NaNO_3PS_2$	5.76	$[M + Na]^+$	251.9895	Pesticide	
Methiocarb sulfone	$C_{11}H_{15}NO_4S$	6.27	$[M + H]^+$	258.0800	Pesticide	
Methiocarb sulfone (Na)*	$C_{11}H_{14}NO_4SNa$	6.27	$[M + Na]^+$	280.0620	Pesticide	
Methiocarb sulfone F1	$C_9H_{12}O_3S$	6.27	Fragment	201.0585	Pesticide	

Compound	Molecular formula	RT (min)	Ion type	Accurate mass	Pollutant family	Source
			ion			
Methiocarb sulfone F2*	C_8H_9O	6.27	Fragment ion	122.0732	Pesticide	
Methiocarb sulfoxide	$C_{11}H_{15}NO_3S$	5.73	[M + H]$^+$	242.0851	Pesticide	
Methiocarb sulfoxide (Na)	$C_{11}H_{14}NO_3SNa$	5.73	[M + Na]$^+$	264.0671	Pesticide	
Methiocarb sulfoxide F1*	$C_9H_{12}O_2S$	5.73	Fragment ion	185.0636	Pesticide	
Thiamethoxam	$C_8H_{10}ClN_5O_3S$	4.26	[M + H]$^+$	292.0271	Pesticide	
Thiamethoxam (Na)*	$C_8H_9ClN_5O_3SNa$	4.26	[M + Na]$^+$	314.0091	Pesticide	
Thiamethoxam F1*	$C_8H_{10}N_4OS$	4.26	Fragment ion	211.0654	Pesticide	
Thiamethoxam F2	C_4H_2NSCl	4.26	Fragment ion	131.9675	Pesticide	
Thiobencarb	$C_{12}H_{16}ClNOS$	13.39	[M + H]$^+$	258.0719	Pesticide	
Thiobencarb (Na)*	$C_{12}H_{15}NaClNOS$	13.39	[M + Na]$^+$	280.0539	Pesticide	
Thiobencarb F1*	C_7H_5Cl	13.39	Fragment ion	125.0158	Pesticide	
Thiodicarb	$C_{10}H_{18}N_4O_4S_3$	9.36	[M + H]$^+$	355.0568	Pesticide	
Thiodicarb (Na)*	$C_{10}H_{17}NaN_4O_4S_3$	9.36	[M + Na]$^+$	377.0388	Pesticide	
Thiodicarb F1	C_3H_5NS	9.36	Fragment ion	88.0221	Pesticide	
Bezafibrate	$C_{19}H_{20}ClNO_4$	11.06	[M + H]$^+$	362.1159	Pharmaceutical	
Bezafibrate (Na)*	$C_{19}H_{19}NaClNO_4$	11.06	[M + Na]$^+$	384.0979	Pharmaceutical	
Chloramphenicol	$C_{11}H_{12}Cl_2N_2O_5$	6.46	[M + H]$^+$	323.0201	Pharmaceutical	
Chloramphenicol (Na)	$C_{11}H_{11}NaCl_2N_2O_5$	6.46	[M + Na]$^+$	345.0021	Pharmaceutical	
Chloramphenicol F1*	$C_{11}H_{10}N_2O_4Cl_2$	6.46	Fragment ion	305.0101	Pharmaceutical	
Chloramphenicol F2*	$C_{10}H_8N_2O_3Cl_2$	6.46	Fragment ion	275.0002	Pharmaceutical	
Chloramphenicol F3	$C_{11}H_8NOCl_2$	6.46	Fragment ion	241.0078	Pharmaceutical	
Gemfibrozil	$C_{15}H_{22}O_3$	13.81	[M + H]$^+$	251.1647	Pharmaceutical	
Gemfibrozil (Na)*	$C_{15}H_{21}NaO_3$	13.81	[M + Na]$^+$	273.1467	Pharmaceutical	

3. Application to samples

After application of the screening strategies to selected food , wastewater and human urine samples, the post-target approach was found to be the most efficient for wide-scope screening. In all samples analyzed, the number of positives was higher than using the non-target approach, in this way giving a more realistic overview of the presence of organic pollutants in the samples. A summary of the results obtained for selected samples is shown in Table 1. The list of pollutants found by target and non-target screening (those with an adequate match reverse fit) is reported together with the main information managed (mass error and retention time), as well as retention time deviation when reference standard was available. Almost in all cases, Rt deviation was lower than 1%. However, the retention time window for positive match was ±0.5 min due to the wide range of matrix analysed having, in some particular cases, deviations higher than 2% typically accepted as in the case of Venlafaxin. Confirmation using MS^E is also shown when it could be made. When reference standards were unavailable, information on fragmentation and retention time was absent. Two strategies were followed to improve the confidence in the compound identification.

The first strategy (Strategy 1 in Table 1) was to simply compare main fragments observed in HE acquisition with common MS/MS product ions reported in the literature for the suspect compound. This was the case for the antibiotic gabapentin, which was detected and identified in urine and wastewater by the presence of two abundant fragments in the HE spectrum with *m/z* 137.0966 and 154.1232 (Fig. 4). These fragment ions were also present in the LE function and had been reported by other authors for determination of gabapentin by QqQ.[43–45] Elemental composition for these two fragments was calculated based on their accurate masses obtaining errors of 0.7 and 0.2 mDa, respectively.

The second strategy (Strategy 2 in Table 1) consisted of justifying the fragments accurate mass (typically observed in the HE spectra) using MassFragment software. This software applies a bond-disconnecting methodology to obtain possible structures for the fragment ions from a given molecule. An example of this approach is shown in Fig. 5, where identification of main fragments of the pharmaceutical irbesartan was carried out. For this purpose, LE and HE combined spectrum of suspect irbesartan was extracted from the chromatographic peak (Fig. 5a and b). The main fragments were justified with the MassFragment tool obtaining reliable structures for all of them. In order to avoid spectrum interferences that could complicate the identification process, recognizing which ions are fragments and which are not, becomes mandatory. From this point of view, UHPLC resolution proved to be valuable for choosing

perfectly coeluting ions (see Fig. 5c). Irbesartan is an angiostensin II receptor antagonist used in the treatment of hypertension that has been in the market for over 10 years.[46] Some fragments observed for irbesartan had been previously reported by ion trap;[46] the most used SRM transition coincides with the most abundant fragment ion of the TOF spectra (*m/z*207.0922).[47,48] However, as MassFragment is a bond-disconnecting software, correct justification is not always feasible.

Thus, for *m/z* 192 ion unreliable structures were suggested. In these cases, previous analyst knowledge or better fragmentation prediction software is necessary.
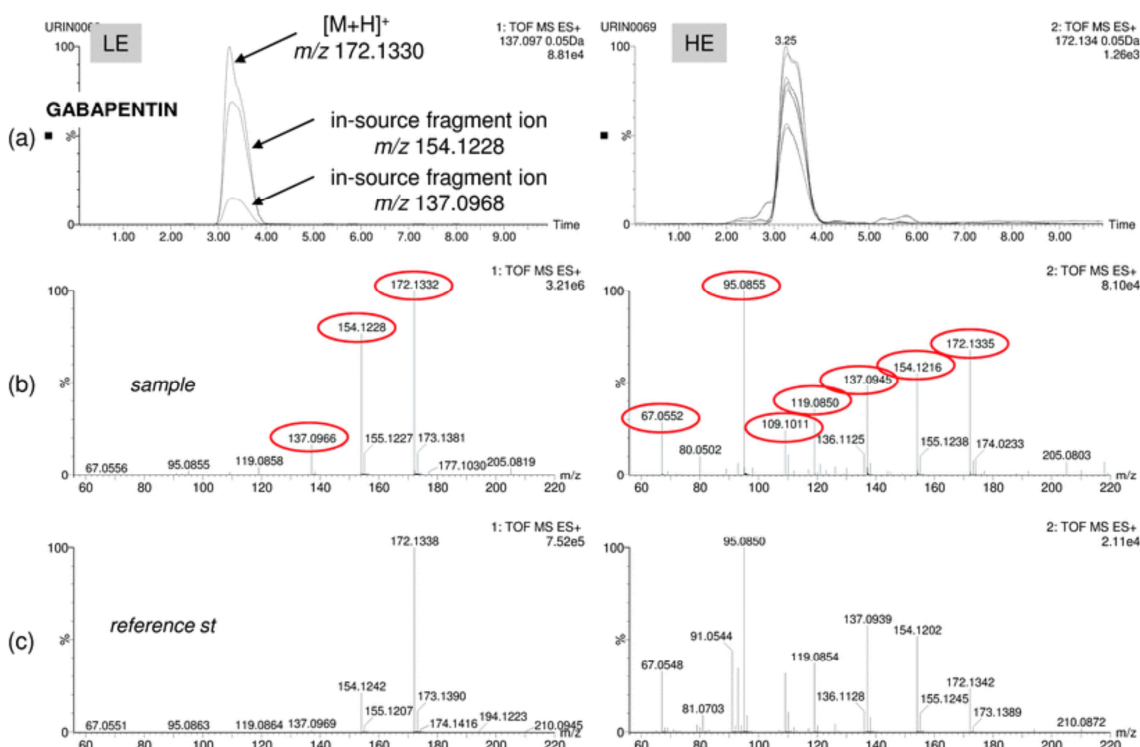
Fig. 4 Positive finding of the pharmaceutical gabapentin in human urine: (a) overlapped nw-XIC for three main ions (protonated ion at *m/z* 172 and in-source fragments at *m/z* 154 and 137) in the LE function and seven coeluting ions in the HE function. LE and HE spectra for sample (b) and reference standard (c) showing good correlation for up to six abundant fragment ions.

In Table 1, the strategy used for the identification of each suspected positive is shown. Bibliographic search and fragment interpretation were helpful to confirm potential positives. When a disagreement occurred between experimental and literature data for fragment ions (if available), and when structures provided by MassFragment software did not fit with the structure

of the candidate, the suspected positive could not be confirmed, and no further research was performed for its elucidation.
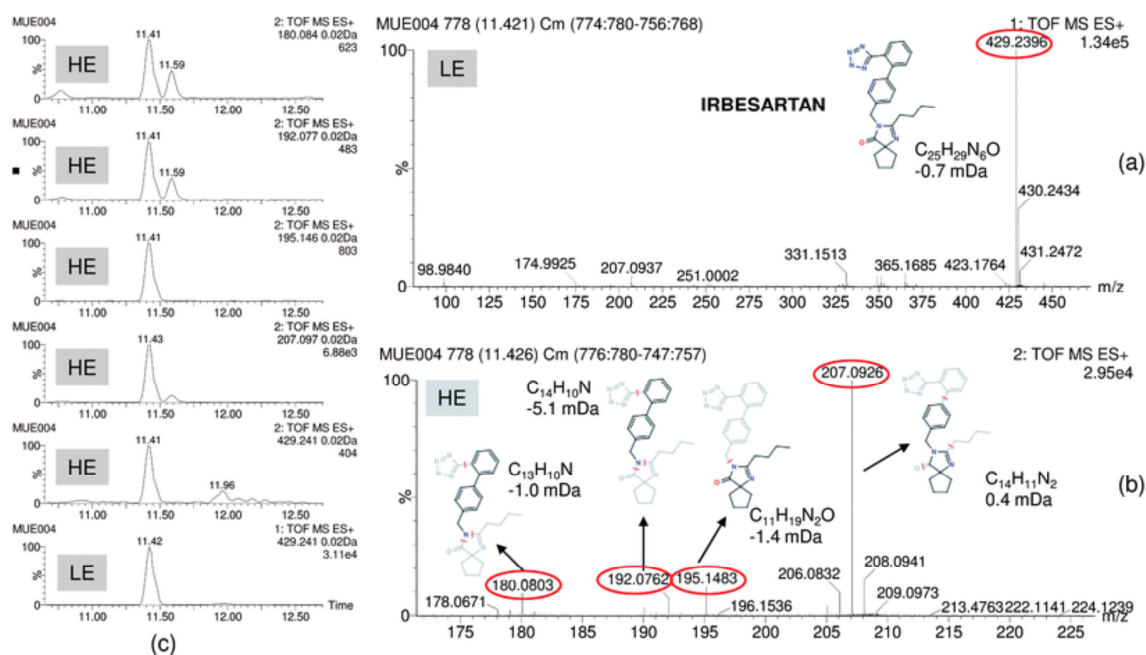
Fig. 5 Positive finding of the pharmaceutical irbesartan in effluent wastewater. Spectra for LE function (a) and HE function (b) of the suspect peak and justification of the HE fragments using MassFragment software. (c) nw-XICs (20 mDa mass window) for [M + H]+ in LE function and main fragments in HE function.

Following the above mentioned strategies, high confidence in the identification process can be achieved. However, no definitive confirmation should be made without injecting the reference standard. Thus, for the most frequently detected pharmaceuticals , irbesartan, valsartan and gabapentin, the reference compounds were acquired. After injecting the standard solutions, all suspect positives in wastewater were confirmed. Our experience on identification of suspect organic contaminants by LC-QTOF under MS^E mode is that the great majority of suspect positives (around 95%) were subsequently confirmed when the reference standard was acquired. This means that acquisition of expensive standards could be made only when solid evidence exists on their presence in samples analyzed. The decision on which

tandards should be acquired would then be made on the basis of previous findings by QTOF MS .
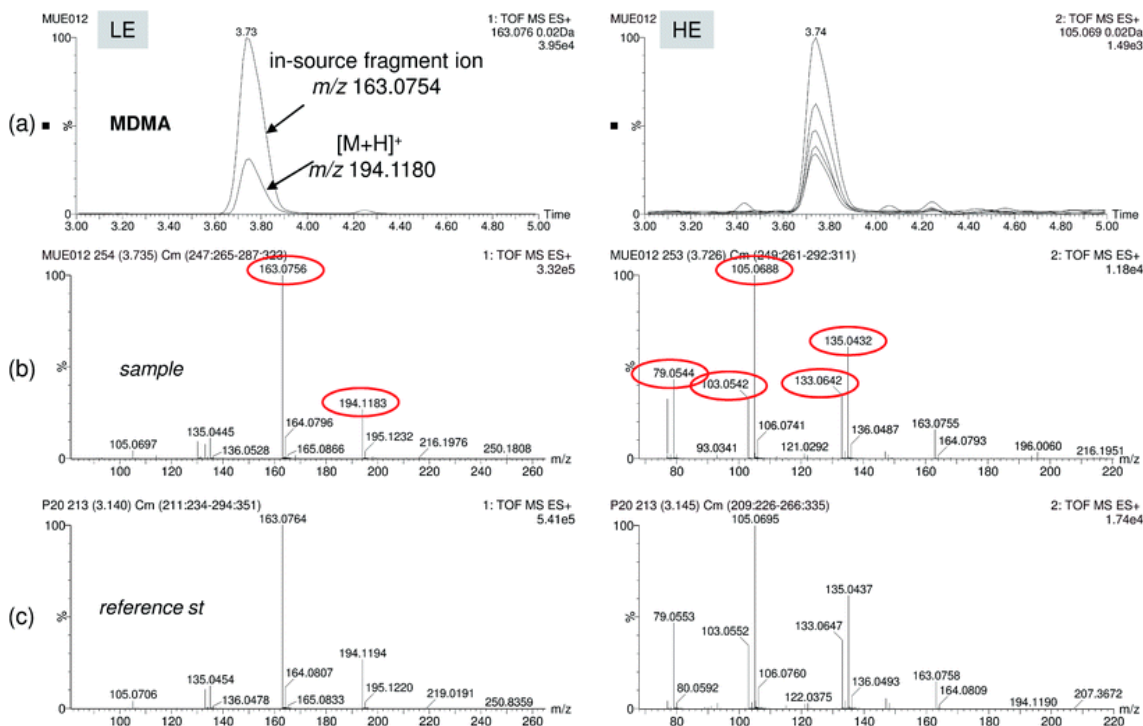
Fig. 6 Positive finding of the drug of abuse MDMA in effluent wastewater. (a) Overlapped nw-XIC for two main ions (protonated ion at *m/z* 194 and in-source fragment at *m/z* 163) in the LE function, and up to five coeluting ions in the HE function. LE and HE spectra for sample (b) and reference standard (c) showing good correlation for up to five abundant fragment ions.

To overcome some post-target limitations and to enhance detectability and identification reliability, improvements in the database approach were made to minimize "missing" compounds due to abundant adduct formation and/or in-source fragmentation. More entries were added in the pollutant database for compounds with a high degree of fragmentation and sodium adduct formation. This is easier when information for the compound is available. After reprocessing the samples using the new, enlarged database, two more compounds (gemfibrozil and bezafibrate) were found in wastewater. These compounds were not detected before due to the abundant sodium adduct formation in positive electrospray ionisation (marked as * in Table 1). In addition, not only the detection step was improved but also the confidence in the identification, as for several analytes, both the protonated molecule and in-source fragments/sodium adducts were also detected (information shown in Table 1). To exemplify this feature, Fig. 6 shows a positive finding of MDMA in EWW. As can be seen, the

protonated ion and main in-source fragment ion of the compound were both detected, the latter being much more abundant than the protonated molecule.

As a summary, two situations could be considered when using the post-target approach based on QTOF measurements:

(a) Detection of target analytes for which standard is available and has been previously injected under the same conditions as the samples. In this case, retention time, in-source fragmentation and adduct formation became useful tools, making the confirmation of findings highly reliable, surely unequivocal.

(b) Detection of suspect compounds for which reference standards are unavailable. Obviously, the situation requires extra-work and time. After a careful study of the full-scan accurate mass data obtained for the suspect compound, a reliable identification could be advanced. A definitive confirmation by injection of the reference standard would be required in the case that significant environmental or legal implications were associated to the presence of the suspect compound. Here, the experience of the analyst and their background on mass spectrometry is of the utmost relevance.

Regarding the non-target screening results, it must be noted that the deconvolution process depends to a great extent on the intensity of the chromatographic peak. Using this approach, several contaminants were missed, as the number of compounds found in the samples was considerably lower than using the targeted one (Table 1). Furthermore, non-target screening with empirical library allowed us to detect very few compounds, not only because of the component detection limitations but also due to fewer entries in this library (231). However, confirmation of the identity becomes simultaneous and more reliable than with other approaches (*i.e.* theoretical library) as LE and HE spectra are compared with those included in the empirical library making unlikely the reporting of false positives. As an example, Fig. 7 shows a corn sample positive tofumonisin . In this figure, two and five coeluting ions were selected for component detection in the LE and HE functions, respectively (Fig. 7a). Both deconvoluted LE and HE mass spectra were automatically compared with those of fumonisin B1 included in the empirical mass spectra library with a match of 80% (Fig. 7b and c).

When employing the theoretical, library-based, non-target screening approach, the investigation of findings when reference standards were unavailable was carried out using the same two strategies discussed before for post-target screening.
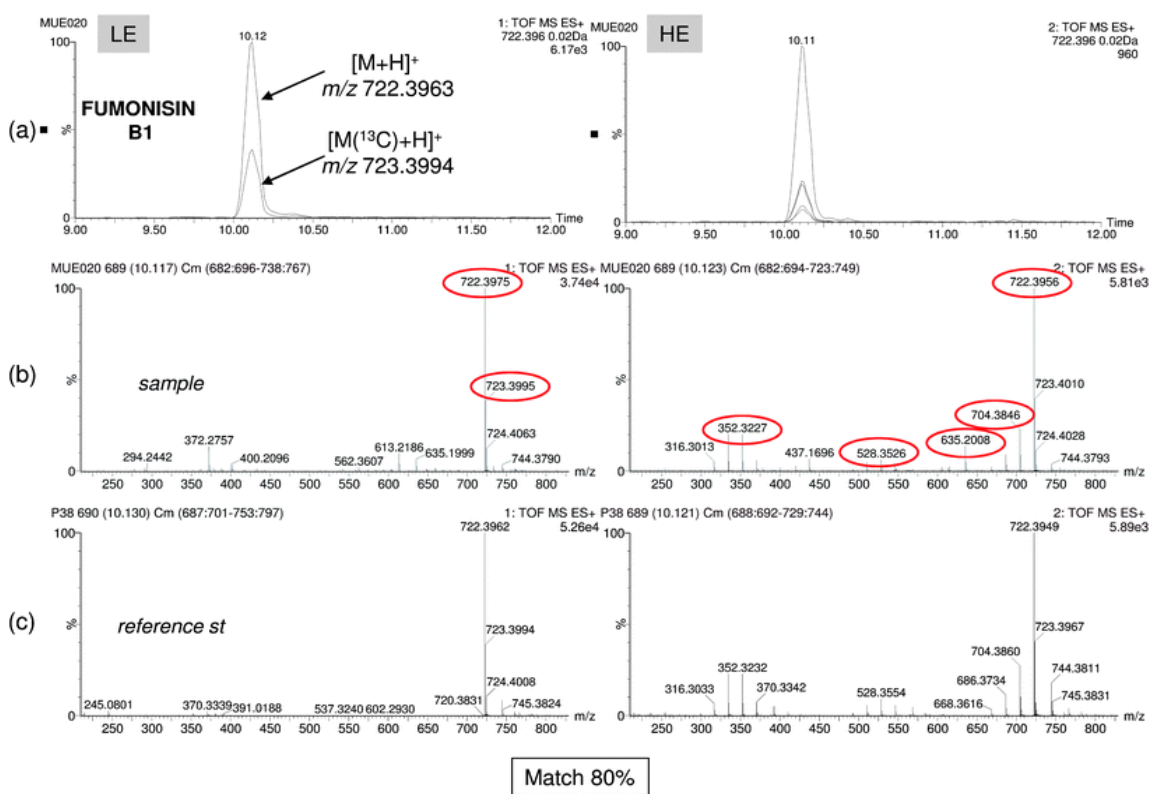
Fig. 7 Non-target screening using experimental library search. Accurate-mass confirmation of the mycotoxinfumonisin B1. (a) Overlapped nw-XICs of the main deconvoluted ions of Fumonisin B1 under LE and HE conditions. Mass spectrum at LE and HE functions for sample (b) and reference standard (c). Library match (80%) and accurate-mass confirmation of the ions (mass errors below 1.3 mDa).

This work shows that the post-target approach has better capability for wide-scope screening of different analyte /sample matrix combinations. However, the non-target approach still has some advantages, especially when using an experimental library, as a comparison of the suspect compound *versus* library spectra is automatically performed achieving a highly reliable identification. In addition, other non-expected compounds that might be present in samples at relatively high concentrations might be detected without any kind of selection (pre- or post-target). However, searching for unknowns is an analytical challenge, where the possibilities to elucidate the components detected are rare.[38] The main limitation for this approach is the difficulty of having large compound libraries similar to those used in GC-MS . At the moment, spectral libraries for LC-MS are home-made and are quite limited. Hopefully, in the near future large standardized libraries which will facilitate non-target screening will be available.

In this work, several contaminants have been found in the three types of samples investigated. Some of them have been tentatively identified without reference standard. The compounds detected belong to very different chemical classes and included emerging contaminants such as pharmaceuticals , UV filters and drugs of abuse, as well as several pesticides . Commonly used post-harvest fungicides imazalil and thiabendazol were identified in the orange and banana samples. In the case of the corn samples, the mycotoxins Fumonisin B1 and B2 were found, and also the less commonly detected Fumonisin B3 and B4 that were not previously included in the common pre-target approaches applied.

CONCLUSIONS

The comparison of different strategies based on the use of UHPLC coupled with QTOF MS for large-scale screening of organic pollutants in food , environmental and urine samples has been carried out. Thanks to the accurate-mass, full-spectrum acquisition in QTOF MS , it is feasible to apply both the target and non-target approaches, which can be seen as complementary within the public health field.

The application of the target approach to selected samples has been demonstrated as an efficient tool for screening a large number of pollutants. For this purpose, a database containing information on the exact mass of the (de)protonated molecule and on the fragment ions and adducts (typically sodium adducts) has been created containing more than 1000 entries. This database has been built on the basis of our own experience and from data reported in the literature on LC-MS analysis of the compounds. Once a compound is detected, the potential positives need to be confirmed taking into account the information obtained on accurate masses of the (de)protonated molecule and of fragment ions, as well as the isotopic distribution. This is feasible using the MS^E acquisition mode in the QTOF instrument, which allows the simultaneous MS data acquisition at low and high collision energy. The accomplishment of retention times and experimental MS^E fragmentation using reference standards obviously facilitates the confirmation step.

In this work, an empirical library containing 231 selected compounds has also been employed in both the target and non-target approaches. Building empirical spectral libraries has been found to be the best way to facilitate both screening types, although it requires the injection of a large number of reference standards to be efficiently applied.

The non-targeted screening presents important drawbacks at low compound concentrations, especially in more complex-matrix samples, due to the difficulties in the

components detection step. Identification of non-target contaminants is greatly facilitated when the compound detected is included in the home-made libraries, otherwise the elucidation of the compound becomes an analytical challenge where the possibilities of success are rare.

An interesting advantage associated with TOF MS -based methodologies concerns the possibility of performing retrospective analysis. This allows investigation of the presence of organic contaminants that were included in the first screening. This can be done at any time, without the need of either new analysis or new sample injections.

ACKNOWLEDGEMENTS

REFERENCES

1   F. Hernández, J. V. Sancho, M. Ibáñez and C. Guerrero, *TrAC, Trends Anal. Chem.*, 2007, **26**, 466

2   A. R. Fernández-Alba and J. F. García-Reyes, *TrAC, Trends Anal. Chem.*, 2008, **27**, 973 .

3   Y. Picó, C. Blasco and G. Font, *Mass Spectrom. Rev.*, 2004, **23**, 45

4   R. Rodil, J. B. Quintana, P. López-Manía, S. Muniategui-Lorenzo and D. Prada-Rodríguez, *Anal. Chem.*, 2008, **80**, 1307 .

5   M. Petrovic, M. Gros and D. Barcelo, *J. Chromatogr., A*, 2006, **1124**, 68

6   E. Beltrán, M. Ibáñez, J. V. Sancho and F. Hernández, *Rapid Commun. Mass Spectrom.*, 2009, **23**, 1801

7   F. Hernández, J. V. Sancho, M. Ibáñez and S. Grimalt, *TrAC, Trends Anal. Chem.*, 2008, **27**, 862

8   A. Kaufmann, P. Butcher, K. Maden and M. Widmer, *J. Chromatogr., A*, 2008, **1194**, 66

9   A. Polettini, R. Gottardo, J. P. Pascali and F. Tagliaro, *Anal. Chem.*, 2008, **80**, 3050

10 I. Ferrer and E. M. Thurman, *J. Chromatogr., A*, 2007, **1175**, 24 CrossRef CAS Search PubMed .

11 A. Kaufmann, P. Butcher, K. Maden and M. Widmer, *Anal. Chim. Acta*, 2007, **586**, 13 .

12 M. J. Martínez Bueno, A. Agüera, M. J. Gómez, M. D. Hernando, J. F. García-Reyes and A. R. Fernández-Alba, *Anal. Chem.*, 2007, **79**, 9372

13 C. A. Mueller, W. Weinmann, S. Dresen, A. Schreiber and M. Gergov, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 1332

14 A. C. Hogenboom, J. A. van Leerdam and P. de Voogt, *J. Chromatogr., A*, 2009, **1216**, 510

15 J. M. Marín, E. Gracia-Lor, J. V. Sancho, F. J. López and F. Hernández, *J. Chromatogr., A*, 2009, **1216**, 1410

16 F. Hernández, O. J. Pozo, J. V. Sancho, L. Bijlsma, M. Barreda and E. Pitarch, *J. Chromatogr., A*, 2006, **1109**, 242

17 B. Kmellár, P. Fodor, L. Pareja, C. Ferrer, M. A. Martínez-Uroz, A. Valverde and A. R. Fernandez-Alba, *J. Chromatogr., A*, 2008, **1215**, 37

18 S. Inoue, T. Saito, H. Mase, Y. Suzuki, K. Takazawa, I. Yamamoto and S. Inokuchi, *J. Pharm. Biomed. Anal.*, 2007, **44**, 258

19 L. Bijlsma, J. V. Sancho, E. Pitarch, M. Ibáñez and F. Hernández, *J. Chromatogr., A*, 2009, **1216**, 3078

20 M. Krauss, H. Singer and J. Hollender, *Anal. Bioanal. Chem.*, 2010, **397**, 943

21 T. Pihlström, G. Blomkvist, P. Friman, U. Pagard and B. Österdahl, *Anal. Bioanal. Chem.*, 2007, **389**, 1773 .

22 C. Lesueur, P. Knittl, M. Gartner, A. Mentler and M. Fuerhacker, *Food Control*, 2008, **19**, 906

23 H. G. J. Mol, P. Plaza-Bolaños, P. Zomer, T. C. De Rijk, A. A. M. Stolker and P. P. J. Mulder, *Anal. Chem.*, 2008, **80**, 9450

24 I. Ferrer, A. Fernandez-Alba, J. A. Zweigenbaum and E. M. Thurman, *Rapid Commun. Mass Spectrom.*, 2006, **20**, 3659

25 Z. Herrera Rivera, E. Oosterink, L. Rietveld, F. Schoutsen and L. Stolker, *Anal. Chim. Acta*, 2011, **70**, 114

26 M. Kellmann, H. Muenster, P. Zomer and H. Mol, *J. Am. Soc. Mass Spectrom.*, 2009, **20**, 1464

27 I. Bobeldijk, J. P. C. Vissers, G. Kearney, H. Major and J. A. Van Leerdam, *J. Chromatogr., A*, 2001, **929**, 63

28 F. Hernández, S. Grimalt, O. J. Pozo and J. V. Sancho, *J. Sep. Sci.*, 2009, **32**, 2245 .

29 F. Hernández, M. Ibáñez, O. J. Pozo and J. V. Sancho, *J. Mass Spectrom.*, 2008, **43**, 173 .

30 S. Grimalt, O. J. Pozo, J. V. Sancho and F. Hernández, *Anal. Chem.*, 2007, **79**, 2833

31 T. Portóles, M. Ibáñez, J. V. Sancho, F. J. López and F. Hernández, *J. Agric. Food Chem.*, 2009, **57**, 4079

32 P. Marquet, N. Venisse, E. Lacassie and G. Lachâtre, *Analysis*, 2000, **28**, 925

33 R. Díaz, M. Ibáñez, J. V. Sancho and F. Hernández, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 355

34 F. Hernández, O. J. Pozo, J. V. Sancho, F. J. López, J. M. Marín and M. Ibáñez, *TrAC, Trends Anal. Chem.*, 2005, **24**, 596

35 M. Ibáñez, J. V. Sancho, F. Hernández, D. McMillan and R. Rao, *TrAC, Trends Anal. Chem.*, 2008, **27**, 481

36 O. J. Pozo, M. Barreda, J. V. Sancho, F. Hernández, J. Lliberia, M. A. Cortés and B. Bagó, *Anal. Bioanal. Chem.*, 2007, **389**, 1765

37 European Commission Decision 2002/657/EC, *Off. J. Eur. Communities: Legis.*, 2002, **221**, 8

38 M. Ibáñez, J. V. Sancho, O. J. Pozo, W. Niessen and F. Hernández, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 169

39 F. Hernández, L. Bijlsma, J. V. Sancho, R. Díaz and M. Ibáñez, *Anal. Chim. Acta*, 2011, **684**, 96

40   T. Portolés, E. Pitarch, F. J. López, J. V. Sancho and F. Hernández, *J. Mass Spectrom.*, 2007, **42**, 1175

41   M. Ibáñez, C. Guerrero, J. V. Sancho and F. Hernández, *J. Chromatogr., A*, 2009, **1216**, 2529

42   O. J. Pozo, P. V. Eenoo, K. Deventer and F. T. Delbeke, *TrAC, Trends Anal. Chem.*, 2008, **27**, 657

43   R. Oertel, N. Arenz, J. Pietsch and W. Kiroh, *J. Sep. Sci.*, 2009, **32**, 238

44   B. Kasprzyk-Hordern, R. M. Dinsdale and A. J. Guwy, *J. Chromatogr., A*, 2007, **1161**, 132 .

45   A. Ojha, R. Rathod, C. Patel and H. Padh, *Chromatographia*, 2007, **66**, 853 .

46   R.   P.   Shah,   A.   Sahu   and   S.   Singh, *J.   Pharm.   Biomed.   Anal.*,   2010, **51**, 1037 CrossRef CAS Search PubMed .

47   G. N. W. Leung, D. K. K. Leung, T. S. M. Wan and C. H. F. Wong, *J. Chromatogr., A*, 2007, **1156**, 271

48   L. Kristoffersen, E. L. Øiestad, M. S. Opdal, M. Krogh, E. Lundanes and A. S. Christophersen, *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.*, 2007, **850**, 147

SUPPORTING INFORMATION

Table 3. Analytes included in the post-target and the non-target with theoretical library screening in **positive ESI mod**e grouped by compounds family.

A) **List of Pesticides included.**

1-Naphtalene acetamide	Cyanofenphos oxygen	Flamprop methyl	Pencycuron
1-Naphthaleneacetic acid	Cycloate	Flazasulfuron	Pendimethalin
2,4,5-T	cycloheximide	Florasulam	Penxonazole
2,4,5-T Isopropyl ester	Cyfluthrin	Fluacrypyrim	Phenmedipham
2,4-D	Cymoxanil	Fluazifop acid	Phenoxyacetic acid
2,4-D Butyl ester	Cyprodinil	Fluazifop-P-butyl	Phenyl mercuric acetate
2,4-D isopropyl ester	Cyproconazole	Flucythrinate	Phorate
2,4-D methyl ester	Cyromazine	Fludioxonil	Phorate oxygen analogue
2-aminobenzimidazol	Dacthal	Flufenacet	Phorate sulfoxide
2-hydroxy-simazine	Daminozide	Flufenoxuron	Phorate-sulfone
2-hydroxy-terbuthylazine	Dazomet	Fluometuron	Phosalone
2-Naphtoxyacetic acid	DDVP (dichlorvos)	Fluquinconazole	Phosfon
2-phenoxypropionic acid	Deethyl ametryn	Fluridone	Phosmet
3,4,5-trimethacarb	Deethyl cyanazine	Fluroxypyr	Phosphamidon
3,4-Dichloraniline	Deethyl cyanazine acid	Flurtamone	Phoxim
5-hydroxy-imidacloprid	Deethyl cyanazine amide	Flusilazole	Picolinafen
5-OH-clethodim-sulfon	Deethyl terbuthylazine	Flutriafol	Pirimicarb
6-chloro-4-hydroxy-3-phenyl-pyridazin	Deethyl-2-hydroxy-terbuthylazine	Folpet	Pirimiphos ethyl
8-hydroxyquinoline	Deethylatrazine	Fonofos	Pirimiphos-methyl
Absidic acid	Deethyldeisopropyl-2-hydroxyatrazine	Forchlorfenuron	Primisulfuron-methyl
ACC	Deethylhydroxyatrazine	Fosmet	Prochloraz
Acephate	Deethylsymetrine	Fosthiazate	Prochloraz
Acetamiprid	Deethylterbumeton	Fuberidazole	Procymidone
Acetate	Deethylterbutryn	Furathiocarb	Profenofos
Acetochlor	Deisopropyl-2-hydroxy-atrazine	Gibberellic acid	Promecarb
Acibenzolar-S-methyl	Deisopropylatrazine	Haloxyfop	Prometon
Aclonifen	Deisopropylprometryne	Haloxyfop-ethoxyethyl	Prometryne
Acrinathrin	Deltamethrin	Haloxyfop-ethoxyethylester	Propachlor
Alachlor	Demethon-S-methyl	Haloxyfop-etotyl	Propamocarb
Alanycarb	Demethon-S-methyl sulfon	Haloxyfop-methyl	Propanil
Albendazole	Demethon-S-methyl sulfoxide	Heptenophos	Propargite

Aldicarb	Demethyl fluometuron	Hexaconazole	Propethamphos
Aldicarb sulfoxide	Demethyl isoproturon	Hexazinone	Propham
Aldicarb sulphone	Demethyl monuron	Hexythiazox	Propiconazole
Ametryne	Desmedipham	Hormodin	Propoxur
Amidosulfuron	Desmethylpirimicarb	Hydroxy atrazine	Propyzamide
Aminocarb	Diafenthiuron	Imazalil	Prosulfuron
Amitraz	Dialifos	Imazameth	Prosulfuron
Anilazine	Diallate	Imazamethabenz-methyl	Prothiophos
Anilofos	Diazinon	Imidacloprid	Pymetrozine
ANTU	Dibenzylamine	Indoxacarb	Pyrazophos
Atrazine	Dichlofluanid	Iprodione	Pyridaben
Avermectin B1a	Dichlone	Iprodione desisopropyl (metabolite)	Pyridaphenthion
Avermectin B1b	Dichlorprop	Iprovalicarb	Pyridate
Azamethiphos	Diclobutrazol	Isazofos	Pyrifenox isomer Z/E 1
Azinphos-ethyl	Diclofop-methyl	Isofenphos	Pyrifenox isomer Z/E 2
Azinphos-methyl	Dicloran	Isoproturon	Pyrimethanil
Azoxystrobin	Dicrotophos	Isoxaflutole	Pyriproxyfen
Benalaxyl	Dicryl	Isoxathion	Quinalphos
Bendiocarb	Dieldrin	Kresoxim-methyl	Quinmerac
Benfuracarb	Diethofencarb	Linuron	Quinoclamine
Bensulfuron-methyl	Difenoconazole	Lufenuron	Quinoxyfen
Bensulide	Difenoxuron	Malaoxon	Quizalofop-ethyl
Bensultap	Diflubenzuron	Malathion	Reserpine
Bentazone	Diflufenican	MCPA	Rimsulfuron
Benzoximate	Disulfoton	MCPA methylester	Rotenone
Bifenazate	Dimefuron	Mecarbam	Silvex
Bifenox	Dimethachlor	Mepanipyrim	Simazine
Bifenthrin	Dimethoate	Merphos	Simetryn
Bitertanol	Dimethomorph estereoisomer 1	Mesotrione	Spinosyn A
Bromacil	Dimethomorph estereoisomer 2	Metalaxyl	Spinosyn D
Bromoxynil	Dimethylvinphos	Metamitron	Spiroxamine
Bromuconazole	Diniconazole	Metazachlor	Strychnine
Bupirimate	Dinocap	Metconazole	Sulfallate
Buprofezin	Diphacinone	Methabenxthiazuron	Sulfosulfuron
Butocarboxym	Diphenylamine	Methamidophos	Sulfotep
Butoxycarboxim-sulfoxid (FC)	Diquat dibromide monohydrate	Methfuroxam	Sulprofos
Butoxycarboxym	Diuron	Methidathion	Tebuconazole
Buturon	Dixanthogen	Methiocarb	Tebufenozide

Cambendazole	Dodemorph	Methiocarb sulfone	Tebufenpyrad
Captan	Dodine	Methiocarb sulfoxide	Tebuthiuron
Carbaryl	Edifenphos	Methomyl	Teflubenzuron
Carbendazim	Endrin	Methoxyfenozide	Terbacil
Carbetamide	EPN	Metobromuron	Terbufos
Carbofuran	Epoxiconazole	Metolachlor	Terbufos sulfone
Carbofuran-3-OH	EPTC	Metosulam	Terbufos sulfoxide
Carbosulfan	Ethephon	Metoxuron	Terbumeton
Carboxin	Ethiofencarb	Metribuzin	Terbuthylazine
Chlorbromuron	Ethiofencarb sulfone	Metsulfuron-methyl	Terbutryn
Chlorfenapyr	Ethiofencarb sulfoxide	Mevinphos 1	Tetrachlorvinphos
Chlorfenvinphos	Ethiofencarbsulfon	Mevinphos 2	Tetraconazole
Chlorfenvinphos-Met	Ethion	Molinate	Thiabendazole
Chlorfluazuron	Ethiprole	Monocrotophos	Thiacloprid
Chloridazon	Ethoxyquin	Monolinuron	Thiamethoxam
Chlorophenoxyacetic acid	Ethofenprox	Monuron	Thifensulfuron-methyl
Chloropicrin	Ethofumesate	Myclobutanil	Thiobencarb
Chlorotoluron	Ethoprophos	Naled	Thiodicarb
Chloroxuron	Etoxazole	Napropamide	Thiofanox
Chlorpropham	Etrimfos	Neburon	Thiofanox-sulfone
Chlorsulfuron	Famoxadone	Nicosulfuron	Thiofanox-sulfoxide
Chromafenozide	Fenamiphos	Nicotine	Thiophanate-ethyl
Cinosulfuron	Fenarimol	N-m-Tolylphthalamic acid	Thiophanate-methyl
Clethodim	Fenazaquin	Nuarimol	Thiram
Clethodim-imin-sulfon	Fenbendazole	Ofurace	Tolclofos-methyl
Clethodim-imin-sulfoxide	Fenbuconazole	Omethoate	Tolylfluanid
Clethodim-sulfon	Fenfuram	Oxamide	Triadimefon
Clethodim-sulfoxid	Fenhexamid	Oxadixyl	Triadimenol
Clodinafop-propargyl	Fenoxaprop	Oxamyl	Triasulfuron
Clofentezine	Fenoxaprop-ethyl	Oxycarboxin	Triazophos
Clomazone	Fenoxycarb	Oxydemeton-methyl	Tribenuron-methyl
Clopyralid	Fenpiclonil	Oxyfluorfen	Trichlorfon
Cloquintocet-mexyl	Fenpropathrin	Oxygen	Tricyclazole
Clorpyrifos	Fenpropidin	Paclobutrazol	Trietazine
Clorpyrifos-methyl	Fenpropimorph	Paraoxon	Trifloxystrobin
Clothianidin	Fenpyroximate	Paraoxon-methyl	Triflumizole
Coroxon	Fensulfothion	Paraquat dichloride	Triflumuron
Coumaphos	Fensulfothion-sulfone	Parathion	Trifluralin
Cyanazine	Fenthion	Parathion-methyl	Trisulfuron-methyl

Cyanazine acid	Fenuron	Parathion-ethyl	Triticonazole
Cyanazine amide	Flamprop	Pebulate	Vamidothion
Cyanofenphos	Flamprop isopropyl	Penconazole	Vinclozolin

B) List of Pharmaceuticals included.

Aceclofenac	Cimaterol	Hydroxymethylclenbuterol	Papaverine
Acetobutolol	Cimbuterol	Ibuprofen	Paracetamol/Acetaminophen
Acetazolamide	Cimetidine	Ibuprofen (Met 1) Carboxyibuprofen	Paroxetine
Acetylsalicylic acid/Aspirine	Ciproheptadina	Ibuprofen (Met 1) Hydroxyibuprofen	Pavabid
Adrafinil	Citalopram	Ifosfamide	Pentoxifylline
Albuterol	Clenbuterol	Indomecthacin	Phenazone/Antipyrine
Aldosterone	Clencyclohexerol	Iopamidol	Phenylbutazone
Alendronic acid	Clenhepterol	Iopromide	Phenylbutazone (Met 1) Oxyphenbutazone
Alprazolam	Clenhexerol	Irbesartan	hydroxylated Phenylbutazone (Met 2)
Amfepramone	Clenisopenterol	Isoprenaline	Phenylthiouracil
Amiloride	Clenpenterol	Isoxsuprine	Pindolol
Aminoglutethimide	Clenproperol	Ketoprofen	Pirbuterol
Amiphenazole	Clobetasol propionate	Lansoprazole	Pravastatin
Anastrozole	Clobetasone butryate	Latanoprost	Prednisolone
Bendroflumethiazide	Clofibrate	Loratadine	Prednisone
Benzylpiperazine	Clofibric acid	Lorazepam	Procaterol
Bumetanide	Clopidogrel	Lorazepam glucuronide	Propanolol
Bupropion	Codeine	Mabuterol	Propylthiouracil
Canrenone	Cortisol/Hydrocortisone	Mahexerol	Propyphenazone
Carphedon	Cortisone	Mapenterol	Quinapril
Carteolol	Cyclophosphamide	Mefenamic acid	Ractopamine
Cathine	Desoxycortisone	Meloxicam	Raloxifene
Celiprolol	Dexamethasone	Mercaptobenzimidazol	Ramipril
Amfetaminil	Diatrizoate	Mestranol	Ranitidine
Chlorothiazide	Diazepam	Metamizole (Met 1) 4-Aminoantipyrine	Reproterol
Chlorphentermine	Diclofenac	Metamizole/Dypirone	Risedronic acid
Chlorthalidone	Diltiazem	Metaprotenerol	Risperidone
Amidotrizoic acid	Dimethylaminophenazone	Methotrexate	Salbutamol
Amitriptyline	Donepezil	Methylprednisolone	Salicylic acid
Amlodipine	Doxazosine	Methylprednisone	Salmeterol
Astemizol	Enalapril	Methylthiouracil	Sertralina

Atenolol	Etofibrate	Metoprolol	Simvastatin
Atorvastatin	Famotidine	Metrapolol	Sotalol
Bamethane	Fenofibric acid/Fenofibrate	Mevastatin	Sulfapyridine
Beclomethasone	Fenoprofen	Mizolastine	Tamsulosin
Benzafibrate	Fenoterol	Montelukast	Tapazol
Benzylthiouracil	Fentanilo	N-(2-Hydroxyethyl)lorazepam	Temazepam
Betaxolol	Florfenicol	Naproxen	Terbutaline
Bethamethasone	Fludrocortisone	Nicotine	Thiouracil
Bezafibrate	Fludrocortisone acetate	Nifedipine	Timolol
Bisoprolol	Flumethasone	N-isopropyl-clenbuterol	Tiotropio bromuro
Brombuterol	Flunixin	Nitroglicerine	Tolfenamic acid
Broom-clenbuterol	Fluocinolone acetonide	N-methyl-mapenterol	Torasemide
Broom-clenproperol	Fluorometholone	Nordiazepam	Tramadol
Broom-mapenterol	Fluoxetine	Noscapine	Triamcinolone
Budesonide	Furosemide	Olanzapine	Triamcinolone acetonide
Buprenorphine	Furosemide glucuronide	Omeprazole	Tulobuterol
Captopril	Gabapentina	Oxazepam	Valsartan
Carbamazepine	Gemfibrozil	Oxprenolol	Venlafaxine
Carbamezapine	Hydrochlorothiazide	Palaudine/(n-hydroxypapaverine)	Warfarin
Carbuterol	Hydrocodone	Pancuronium	Zafirlukast
Carprofen	Hydroxy-clenbuterol	Pancuronium Bromide	Zilpaterol
Chloramphenicol	Hydroxy-clenhepterol	Pantoprazol	
AHD	Dicloxacillin	Midecamycin	Spiramycin II
Albendazole	Difloxacin	Minocycline	Spiramycin III
Amikacin	Dimetridazole	Monesin	Streptomycin
Amoxicillin	Doxycycline	Moxifloxacin	Sulfacetamide
AMOZ	Doxytetracycline	Nafcilline	Sulfachlorpyrazine
Ampicilline	emamectin B1a	Nalidixic acid	Sulfachlorpyridazine
AOZ	emamectin B1a benzoate	Narasin	Sulfadiazine
Azaperol	emamectin B1b	Naxilinic acid	Sulfadimethoxine
Azaperone	emamectin B1b benzoate	Nitrofurantoin	Sulfadimidine
Azepromazine	Enoxacin	Nitrofurazone	Sulfadoxine
Azithromycin (+2)	Enrofloxacin	Norfloxacin	Sulfaguanidine
Carozolole	Erythromycin (-H2O)	Ofloxacin	Sulfamerazine
Cefaclor	Erythromycin A	Oleandomycin	Sulfamethazine
Cefazoline	Erythromycin B	Cyclamycin	Sulfamethizole
Cefoperazone	Erythromycin C	Oxacillin	Sulfamethoxazole
Cefotaxime	Ethopabate	Oxfendazole	Sulfamethoxypyridazine

Cefoxatime	Febantel	Oxibendazole	Sulfamonometoxine
Ceftazidime (2+)	Fenbendazole	Oxonilic acid	Sulfanilamide
Ceftizoxime	Fleroxacin	Oxytetracycline	Sulfapyridine
Ceftriaxone	Flubendazole	Pefloxacin	Sulfaquinoxaline
Cephadroxile	Flumequine	Penicillin	Sulfasalazine
Cephalexin	Furaltadone	Penicillin G	Sulfathiazole
Cephapirin	Furazolidone	Penicillin V	Sulfisoxazole
Cephradine	Gatifloxacin	Penicillines generic fragment	Sulfonamides generic fragment
Chlorpromazine	Gemifloxacin	Pipedimico	Temafloxacin
Chlortetracyclin	Gentamicin	Piperacillin	Tetracycline
Cinoxacin	Grepafloxacin	Pirlimycin	Tiamulin
Ciprofloxacin	Griseofulvin	Piromidic acid	Tilmicosin
Ciprofloxacin (Met 1)	Halofuginone	Propionylpromazine	Tinidazole
Ciprofloxacin (Met 2)	HMMNI	Rifampicin	Tobramycin
Ciprofloxacin (Met 3)	Josamycin	Robenidine	Triclabendazole
Ciprofloxacin (Met 4)	Lasalocid	Ronidazole	Trimethoprim
Ciprofloxacin (Met 5)	Leucomalachite green	Rosoxacin	Trovafloxacin
Clarythromycin	Levamisole	Roxythromycin	Tylosin A
Clindamycin	Lincomycin	Rufloxacin	Tylosin B
Clopidol	Lomefloxacin	Salinomycin	Tylosin C
Cloxacillin	Maduramicin	Sarafloxacin	Tylosin D
Danofloxacin	Malachite green	SEM	Xylazine
Diaveridine	Marbofloxacin	Sparfloxacin	
Diclazuril	Metronidazole	Spiramycin I	

C) List of Illicit drugs included.

10-Hydroxymorphine	Cocaine	LSD	Nor LSD / nor-iso-LSD
11-hydroxy-THC	Codeine	MDA	Norbenzoylecgonine
11-nor-9-carboxy-Δ9-THC	Codeine epoxide	MDEA	Norcocaethylene
14-Hydroxycodeine	Codeine-6-glucuronide	MDMA	Norcocaine
2-oxo-3-hydroxyLSD	Cotinine	Methadone	Norcodeine
3-Acetylmorphine	Diacetylmorphine	Methamphetamine	Normorphine
3'-Hydroxycocaine	Ecgonidine	Methylecgonidine	Normorphine-6-glucuronide
3-Hydroxyflunitrazepam	Ecgonine	Methylecgonine	Nortropacocaine
4-hydroxybutanoate	Ecgonine methyl esther	m-Hydroxybenzoylecgonine	Paramorphine
6-Acetylmorphine (6-AM)	EDDP	Morphine	Paraxanthine
7-aminoflunitrazepam	Ephedrine	Morphine-3β-glucuronide	PCP

7-aminonorflumitrazepam	Ethylecgonine	Morphine-6β-glucuronide	Pseudomorphine
Acetylcodeine	Ethylen cocaine	Morphine-7,8-oxide	Pseudomorphine (C17 alkaoid)
Amphetamine	Fentanyl	Morphine-gsh	Psicolin
Benzoylecgonine	Flunitrazepam	N-Benzoylnormethylecgonine	Speed
Caffeine	GHB	N-Demthylflunitrazepam	THC
Cinnamoylcocaine	Heroin	N-Hydroxynorcocaine	THC-glucoronide
Cocaethylene	Ketamine	Nicotine	

D) List of Hormones included.

1,4,6-androstatrien-17-ol-3-one	Androstanolone	Ethinylestradiol	Nandrolone
1,4-androstadiene-3,17-dione	Bolandiol	Ethisterone	Nandrolone
16B-hydroxyfurazabol	Bolasterone	Ethylestrenol	Norclostebol
16B-hydroxystanozolol	Boldenone	Ethynodiol	Norclostebol
17alpha-Trenbolone	Boldenone metabolite	Ethynyloestradiol	Norethandrolone
17B-oestradiol	Broparoestrole	Fluoxymesterone	Norethisterone
17-hydroxy-progesterone	Calusterone	Formebolone	Norethynodrel
17-oestradiol	Chlormadinone	Formebolone metabolite	Norprogesterone
17-Methyl-1.testosterone	Chlormadinone acetate	Gestonorone-caproate	Norvinisterone
19-hydroxyandrostenedione	Clostebol	Gestrinone	Oestriol
19-Nortestosterone	Cyclofenil	Hexestrol	Oestrone
1-testosterone	Cyproteron	Lynestrol	Oxandrolone
3'-hydroxy-progesterone	Danazol	Medrogestone	Oxymesterone
3'-hydroxystanozolol	Dehydrochloromethyl testosterone	Medroxyprogesterone	Oxymetholone
4,6-androstadien-17-ol-3-one	Delmadinone acetate	Medroxyprogesterone acetate	Progesterone
4,9-androstadien-17-methyl-17-ol-3-one	DHT	Megestrol acetate	Quinbolone
4,9-androstadien-17-ol-3-one	Dienestrol	Melengestrol	Quinestrol
4B-hydroxy-progesterone	Diethylstilbestrol	Melengestrol acetate	Stanozolol
4-Cl-17a-methyl-androsta-1,4-diene-6b,17b-diol-3-one	Dimethisterone	Mestanolone	Stenbolone
4-hydroxystanozolol	Dl-norgestrel	Mesterolone	Taleranol
4-hydroxytestosterone	Drostanolone	Mestranol	Tamoxifen
6alpha-hydroxy-androst-4-ene-3,17-dione	Dydrogesterone	Methandienone	Testosterone
6b-hydroxymetandienone	Epimestrol	Metenolone	THG
6b-hydroxytestosterone	Epi-nortestosterone	Methandrol	Trenbolone
6-hydroxy-4-Cl-dehydromethyltestosterone	Epi-testosterone	Methenolone	Vinyltestosterone
6-hydroxyandrostenedione	Epi-trenbolone	Methylestrenolone	Zearalenone

Algestone acetophenide	Estradiol	Methyltestosterone	Zearanol
Allylestrenol	Estriol	Methyltrienolone	Zearalenol
Allyltrenbolone	Estrone	Mibolerone	

E) List of Mycotoxins included.

15-Acetoxyscirpenol	Beta-Zearalenol	Ergosine	Moniliformin Sodiium
15-Acetyldeoxynivalenol	Chaetoglobosin A	Ergotamine	Monoacetoxyscirpenol
3-Acetyldeoxynivalenol	Chanoclavine	Festuclavine	Mycophenolic acid
3alpha-Acetyldiacetoxyscirpenol	Chetomin	Fumitremorgin C	Neosolaniol
5-Methoxysterigmatocystin	Citrinin	Fumonisin A1	Nivalenol
Aflatoxicol I	Cytochalasin E	Fumonisin A2	Norsolorinic acid
Aflatoxin B1	Deepoxy-deoxynivalenol	Fumonisin B1	Ochratoxin Alpha
Aflatoxin B2	Deoxynivalenol	Fumonisin B2	Ochratoxin B
Aflatoxin B2 alpha	Deoxynivalenol-3-glucoside	Fumonisin B3	Parasiticol
Aflatoxin G1	Diacetoxyscirpenol	Fumonisin B4	Paxilline
Aflatoxin G2	Dihydroergosine	Fusarenon-X	Penitrem A
Aflatoxin M1	Dihydroergotamine	Gibberallic acid	Roridin A
Aflatoxin M2	Dihydrolysergol	Gliotoxin	Scirpentriol
Agroclavine	Dihydroxysterigmatocystin	Griseofulvin	Sterigmatocystin
Alpha-ergocryptine	Elymoclavine	HT-2 Toxin	Sulochrin
Alpha-Zearalenol	Emodin	Hydrolyzed zearalenone	T-2 Toxin
Altenuene	Enniatin A	Lysergol	T-2 Triol
Alternariol	Enniatin B	Meleagrin	Tentoxin
alternariol methyl ether	Ergine	Methis-ergide	Verrucarin A
Austocystin A	Ergocornine	Methoxysterimatocystin	Verrucarol
Avenacein Y	Ergocristine	Mevinolin	Verruculogen
Averufin	Ergometrine	Moniliformin	Zearalenone
Beauvericin	Ergometrinine	Moniliformin Potassium	Zearalenone-4-glucoside

F) List of Surfactants included.

C11-DEA	C12-AEO9	C14-AEO8	C16-AEO9
C12-AEO10	C13-DEA	C14-AEO9	C18-AEO10
C12-AEO11	C14-AEO10	C15-DEA	C18-AEO3
C12-AEO12	C14-AEO11	C16-AEO10	C18-AEO4
C12-AEO13	C14-AEO12	C16-AEO11	C18-AEO5
C12-AEO14	C14-AEO13	C16-AEO12	C18-AEO6
C12-AEO3	C14-AEO14	C16-AEO4	C18-AEO7

C12-AEO4	C14-AEO4	C16-AEO5	C18-AEO8
C12-AEO5	C14-AEO5	C16-AEO6	C18-AEO9
C12-AEO6	C14-AEO6	C16-AEO7	C7-DEA
C12-AEO7	C14-AEO7	C16-AEO8	C9-DEA

G) List of Dyes included.

Brilliant green	Sudan Black B	Sudan Orange G
Butter yellow	Sudan I	Sudan Para-red
Crystal Violet	Sudan II	Sudan Red 7B
Leucocrystal violet	Sudan III	Sudan Red B
Rhodamine B	Sudan IV	Sudan Red G

H) List of UV filters included.

| 4-MBC | BP-3 | OC |
| BM-DBM | IAMC | OD-PABA |

I) List of Phenols included.

2,4-dimethylphenol	4-chloro-3-methylphenol	Catechol	o-Cresol
Chlorophenol	4-chlorophenol	Dinoseb	p-Cresol
3-methyl-4-nitrophenol	Bisphenol A	m-Cresol	Phenol

SUPPLEMENTARY INFORMATION

Table 4. Analytes included in the post-target and the non-target with theoretical library screening in **negative ESI mode** grouped by compounds family.

A) List of Pesticides included.

2,4-D	Chlorsulfuron F1	Fluroxypyr F2*	Primisulfuron methyl
2,4-DB	Dicamba	Haloxyfop	Propanil
6-chloronicotinic acid	Dichlorprop	Hexaflumuron	Teflubenzuron
Absidic acid	Fenazaquin	Ioxynil	Terbacil
Acifluorfen	Fenpropathrin	MCPA	Toxaphene
Bentazone	Fludioxonyl	Mecoprop	Triflumuron
Bromacil	Fludiozonil	Mestrione	
Bromoxynil	Fluroxypyr	Oxyfluorfen	
Chlorsulfuron	Fluroxypyr F1	Pentachlorophenol	

B) List of Pharmaceuticals included.

Acetazolamide	Fenoprofen	Cefuroxime
Acetylsalicylic acid/Aspirine	Furosemide	Chloramphenicol
Adrafinil	Ibuprofen	Diclazuril
Bendroflumethiazide	Ketoprofen	Enrofloxacin
Bezafibrate	Mefenamic acid	Flumequine
Clofibric acid	Naproxen	Nalidixic acid
Diclofenac	Nimesulide	Nicarbazine
diclofenac sodium salt	Piroxicam	Nitrofurantoina
Diflusinal	Pravastatin	Oxonilic acid
Digoxin	Salicylic acid	Sulfamethoxazole
Enalapril	Sulfasalazine	
Enalaprilat	Valsartan	

C) List of Illicit drugs included.

4-hydroxybutanoate
11-nor-9-carboxy-Δ9-THC
GHB

D) List of Hormones included.

Boldenone	Progesterone	Trenbolone
Estrone	Testosterone	Zearalenol
Nandrolone		

E) List of Mycotoxins included.

Citrinin	Mycophenolic acid
Gibberallic acid	Patulin

F) List of Surfactants included.

C10-LAS	C12-LAS	C14-AE1S	NPEO2-SO4
C11-LAS	C13-AE1S	C14-AE2S	NPEO3-SO4
C12-AE1S	C13-AE2S	C14-AE3S	NPEO4-SO4
C12-AE2S	C13-AE3S	C14-AE4S	NPEO5-SO4
C12-AE3S	C13-AE4S	C14-AE5S	NPEO6-SO4
C12-AE4S	C13-AE5S	C14-AS	NPEO7-SO4

C12-AE5S	C13-AE6S	C16-AS	OP1EC
C12-AE6S	C13-AS	DHSS	
C12-AS	C13-LAS	NP1EC	

G) List of Dyes included.

Acid blue 1	Brilliant blue FCF	Metanil Yellow	Quinoline yellow
Acid green 9	Chrysoine	Naphtol yellow S	Red 10B
Acid red 52	Eosine	New coccine	Red 2G
Acid violet	Erythrosine	Orange I/II/RN	Rose bengal
Allura Red AC	Fast green FCF	Patent blue V	Sunset Yellow
Amaranth	Fast Red E	Phloxine B	Tartrazine
Amido black 10B	Fast Yellow AB	Ponceau 3R	Uranine
Azorubine	Green S	Ponceau 4R	Yellow 2G
Benzyl violet 4B	Indigotine	Ponceau 6R	
Brilliant black PN	Martius yellow	Ponceau SX	

H) List of UV filters included.

Benzophenone-1 (BP-1)	PDT	2,2'-dihydroxy-4-methoxy benzophenone
Benzophenone-2 (BP-2)	4-hydroxybenzophenone (HBP)	2,3,4-Trihydroxybenzophenone (THB)
Benzophenone-3 (BP-3)	4-hydroxybenzophenone (HBP)	Terephthalylidene dicamphor sulfonic acid
Benzophenone-4 (BP-4)	2-hydroxy-4-methoxy benzophenone	Ethylhexyl salicylate (EHS)
PBSA	2,4-dihydroxy benzophenone (DHB)	

I) List of Phenols included.

2,3,4,5-tetrachlorophenol	2,4-diclorophenol	3,4-dichlorophenol	4-octylphenol
2,3,4,6-tetrachlorophenol	2,4-dimethylphenol	3,5-dichlorophenol	Bisphenol A
2,3,4-trichlorophenol	2,4-dinitrophenol	3,5-dinitrophenol	Catechol
2,3,5,6-tetrachlorophenol	2,5-dichlorophenol	3-chlorophenol	Dinoseb
2,3,5-trichlorophenol	2,6-dichlorophenol	3-methyl-4-nitrophenol	m-Cresol
2,3,6-trichlorophenol	2,6-dinitro-2-methylphenol	4,6-dinitro-4-methylphenol	o-Cresol
2,3-dichlorophenol	2-chlorophenol	4.nonylphenol	p-Cresol
2,4,5-trichlorophenol	2-nitrophenol	4-chloro-3-methylphenol	Pentachlorophenol
2,4,6-trichlorophenol	p-nitrophenol	4-chlorophenol	Phenol
2,4,6-trichlorophenol	3,4,5-trichlorophenol	4-nitrophenol	

2.3.1. Discusión de resultados

Este trabajo representa la continuación del primer artículo científico en el cual se estudiaban las condiciones óptimas para la creación de una librería de espectros y el posterior análisis de tipo screening por LC-HRMS. En él se comparan diferentes aproximaciones para afrontar el tratamiento de la ingente cantidad de datos obtenidos en base a la automatización, confianza de la identificación, al número de falsos positivos y, sobretodo, al de falsos negativos. En un screening es importante poder asegurar, con la máxima fiabilidad posible, si una muestra está o no contaminada a niveles de concentración relativamente bajos previamente establecidos. Por ello, el procesamiento de datos no debe presentar un impacto negativo sobre la detección/identificación de los diferentes compuestos, pues restaría robustez a la metodología aplicada. Se estudiaron dos aproximaciones: non-target y post-target. El término post-target, ha sido comentado en el capítulo de introducción, y se refiere a la búsqueda o selección de los analitos después de la inyección de la muestra y de haber adquirido los datos. Esta fue la opción más adecuada, gracias en parte al bajo número de falsos positivos y negativos encontrados.
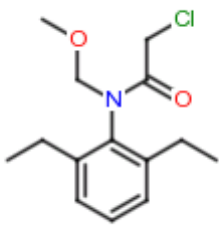
Así pues, mediante la metodología post-target fue posible detectar un mayor número de analitos en comparación a las aproximaciones non-target, tanto si se utilizaba la librería de espectros doméstica experimental o teórica. Sin embargo, en la aproximación non-target con librería empírica, la confianza en la identificación de los compuestos encontrados fue mayor gracias a la comparación automática con los espectros de los patrones incluidos en la librería tanto a baja como a alta energía de colisión. Esto permitía de manera sencilla y visual (además de por los valores de similitud/*match*) confirmación inmediata de la identidad. Para mejorar las posibilidades identificativas en la aproximación post-target se introdujeron los datos de los fragmentos observados para los compuestos incluidos en la librería experimental (por tanto, se disponía de patrón), como entradas adicionales de la base de datos utilizada para el screening post-target. Para ello, se calcularon las composiciones elementales de cada uno de los fragmentos de todos los compuestos mediante el uso del software MassFragment.

El programa MassFragment se basa en cálculos de rotura de enlaces con el fin de encontrar subestructuras posibles para cada uno de los iones de un espectro concreto a partir de la estructura química del compuesto en cuestión. Esto significa que a partir de una estructura y de los iones en masa exacta observados en el espectro de fragmentación, el software calcula automáticamente donde podría producirse la rotura del enlace o enlaces para obtener dichos fragmentos. Además, puede seleccionarse, a elección del usuario, un margen de error en cuanto a los hidrógenos de la estructura del fragmento final, permitiendo así predecir la posible (des)protonación, ciclación o formación de dobles enlaces durante el proceso de fragmentación. Adicionalmente, se puntuan las posibles estructuras propuestas para los posibles fragmentos en base al tipo de enlace/s roto/s para la formación de ésta, es decir, de entre las diferentes subestructuras posibles para un determinado fragmento, se penalizan en mayor medida aquellas que involucren roturas de enlace de grupos aromáticos, ciclos o enlaces múltiples, pues este tipo de fragmentaciones son energéticamente menos probables y, por tanto, menos comunes.

En la figura 2.B. se muestra un ejemplo de los resultados obtenidos para el pesticida Alachlor. En este caso, el compuesto se encontraba altamente fragmentado incluso en la primera función (baja energía de colisión, LE) y además presentaba un aducto sodiado abundante, siendo prácticamente inexistente el ion correspondiente a la molécula protonada. Además, se puede observar, de forma clara el *pattern* isotópico correspondiente a un átomo de Cl en uno de los fragmentos, lo que facilitó en gran medida la elucidación de los fragmentos. Sin embargo, esta aproximación presenta dos principales inconvenientes. El primero de ellos, puede observarse en la figura 2.B., donde el *clúster* correspondiente al aducto sodiado no es justificado debido a que el programa no está diseñado para deducir iones de mayor masa que la molécula en cuestión, no pudiendo predecir la formación de aductos o de dímeros. El segundo viene dado por la incapacidad de deducción de fragmentos formados a partir de reordenamientos en la molécula, ya que únicamente se consideran roturas de enlace, o posibles formaciones de iones policargados (aunque esto último es muy improbable en el caso de moléculas pequeñas). Estos inconvenientes deben pues ser solventados por el analista a partir de su conocimiento y experiencia.
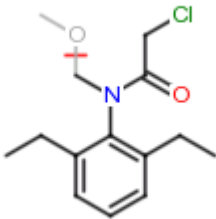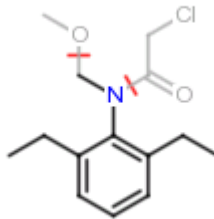
Alachlor

Input:

	ID (job)	9
	Mass (Da)	269.1183
	Formula	$C_{14}H_{20}NO_2Cl$
	DBE	5

Experiment:

Product ion(s) (Da)	162.1282 238.1005 239.1046 239.1635 240.0978 240.1678 292.1091 293.1147 294.1065 295.1085 +/- 0.01 in positive mode, structure filter off
DBE	-1.5 to 50
Electron count	both
Maximum H deficit	6
Fragment number of bonds	4
Scoring	aromatic: 6, multiple: 4, ring: 2, phenyl: 8, other: 1 H-deficit: 0, hetero modifier: 0.5, max score: 16

Results:

238.1005 ¬+ (+0H)	162.1282 ¬+ (+1H)
238.0999 (+0.6.mDa) (S:0.5, B:1)	162.1283 (-0.1.mDa) (S:1.0, B:2)

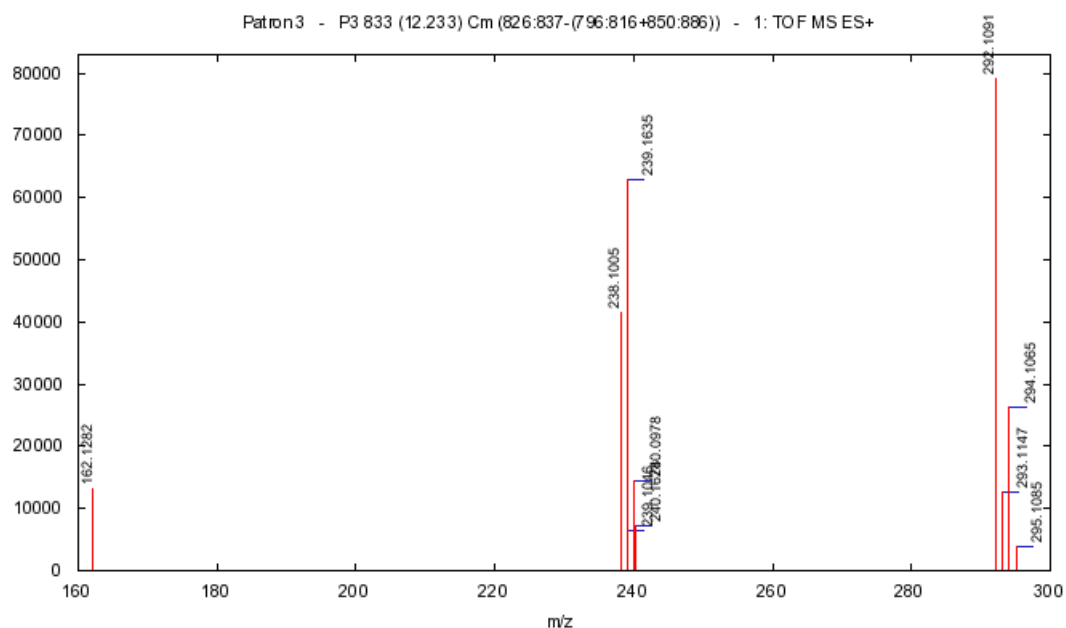| $C_{13}H_{17}NOCl$ (-CH$_4$O)	$C_{11}H_{16}N$ (-C$_3$H$_5$O$_2$Cl)

Figura 2.B. Resultados obtenidos mediante MassFragment para la elucidación estructural del pesticida Alachlor a partir del espectro del patrón correspondiente en la función de baja energía de colisión.

La metodología aplicada permitió la detección e identificación de compuestos cuyos tiempos de retención y fragmentos fueran previamente conocidos (es decir cuyo patrón estuviese disponible y hubiese sido previamente inyectado), pero, además, estaba disponible la información de fragmentación de los compuestos incluidos en la base de datos teórica (en el caso de detectarlos en la muestras), de los cuales solo se tenía información de la fórmula molecular. Esta información permitió la identificación tentativa de aquellos compuestos cuya presencia fuera reiterada, como por ejemplo el fármaco Irbesartan, mediante el uso de los espectros de baja y alta energía y la justificación de fragmentos usando MassFragment o la búsqueda en literatura. La experiencia en cuanto a esta pseudo-confirmación, o identifiación tentativa, nos ha demostrado que casi la totalidad de los positivos encontrados y

razonadamente justificados han sido posteriormente confirmados mediante la inyección de patrones bajo las condiciones del método. Un ejemplo de este procedimiento se muestra en la figura 5 del artículo.

La identificación tentativa de compuestos sin patrón permitió ampliar el conocimiento sobre la presencia de contaminantes emergentes en las muestras analizadas, e incluir posteriormente dichos compuestos en métodos multiresiduales de LC-QqQ MS para su control en muestras medioambientales (Gracia-Lor, Martínez et al. 2012). Nuestros estudios con QTOF MS revelaron la presencia de fármacos que no habían sido incluidos en la base de datos, como el antes mencionado Irbesartan, en un importante número de muestras de agua analizadas.

Este segundo artículo se sentaron las bases para una metodología de screening "universal", que con el fin de reforzarla debió ser validada cualitativamente, tal como se muestra en el siguiente artículo científico.

2.4. Artículo científico 3

"Qualitative validation of a liquid chromatography-quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters"

R. Díaz, M. Ibáñez, J.V. Sancho and F. Hernández.

Journal of Chromatography A, 1276 (2013) 47-57.

Qualitative validation of a liquid chromatography–quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters

R. Diaz, M. Ibáñez, J.V. Sancho, F. Hernández*

Research Institute for Pesticides and Water, University Jaume I, 12071 Castellón, Spain

ABSTRACT

A multiclass wide-scope screening of organic contaminants in natural and waste water has been developed and validated for qualitative purposes, i.e. detection and reliable identification of compounds detected in samples at a certain level of concentration. The screening is based on the use of liquid chromatography coupled to quadrupole-time of flight mass spectrometry (LC–QTOF MS) and has been applied to water samples of different origin and matrix composition (surface water, ground water and effluent urban wastewater). Water samples were spiked with a standard mixture of around 150 organic contaminants from different chemical families (including a number of relevant metabolites/transformation products (TPs), at 0.1 and 1 μg/L concentration levels. After solid-phase extraction with Oasis HLB cartridges, sample extracts were analyzed by LC–QTOF MS and the accurate-mass full-spectrum data were processed for qualitative analysis. The presence of at least two ions (typically the (de)protonated molecule and one fragment ion) accurate-mass measured was used for the reliable identification. The screening detection limit (SDL) and the limit of identification (LOI) were established as the main parameters of the screening method. Nearly all compounds could be detected at the lowest concentration tested, but identification was problematic for some compounds at 0.1 μg/L level, especially in wastewater samples. The screening procedure was finally applied to different water samples using a home-made database of around 1100 organic contaminants. It allowed the detection and identification of several antibiotics, anti-inflammatory/analgesics drugs and lipid regulators. Cocaine and its metabolite benzoylecgonine were also frequently detected. In addition, triazine herbicides and their TPs, and fungicides like thiabendazol, carbendazim or imazalil, were also identified in some of the samples.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Environmental pollution is a matter of increasing concern as the number of contaminants potentially hazardous for plant and animal species, or even for human health, is extraordinary high. During many years, the impact of chemical pollution has been focused almost exclusively on the so-called "priority" pollutants, especially those acutely toxic/carcinogenic pesticides and industrial intermediates displaying persistence in the environment. However, in the last years, the presence of emerging contaminants, such as pharmaceuticals [1–3], illicit drugs [4–6] or personal care products [7–9], is being reported in the environment. These compounds can finally reach the aquatic environment due to the inability of water treatment plants to completely remove some of them [10].

Environmental analytical chemists have to face the investigation of thousands of organic pollutants from different origin and chemical composition that may be present in the aquatic environment. Multiclass/multiresidue methods, commonly based on LC–MS/MS, are applied to monitor most emerging contaminants, as this technique allows the satisfactory identification and quantification of target analytes at low concentrations. However, although quantitative data are highly useful, information on the possible presence of other pollutants that do not form a part of the list is missed in most target methods [11–13].

Full-spectrum acquisition techniques based on the use of high resolution mass spectrometry (HRMS) are increasingly being applied for wide-scope screening purposes in the environmental field [12,14,15]. Screening a large number of contaminants is highly useful in the environment to have realistic information of the actual composition of samples [16]. HRMS can achieve extended multi-class compounds information as well as provide a broad and realistic knowledge about sample pollution in a rapid, sensitive and selective way [17]. Moreover, this approach facilitates the easy method update or even the retrospective analysis of suspected samples in a subsequent step, a relevant fact as "emerging contaminants" are continuously appearing (as well as new transformation products discovered), this being another new reason of concern [18].

* Corresponding author. Tel.: +34 964387366; fax: +34 964387368.
E-mail address: felix.hernandez@uji.es (F. Hernández).

Qualitative validation of a LC-(Q)TOF MS screening method for organic pollutants in waters

R. Díaz, M. Ibáñez, J.V. Sancho and F. Hernández

Research Institute for Pesticides and Water, Universitat Jaume I, E-12071 Castelló, Spain

HIGHLIGHTS

• Qualitative validation of a UHPLC–(Q)TOF MS wide-scope screening method performed.
• Contaminant identification criteria have been discussed for HRMS methodologies.
• SDL and LOI have been established for around 150 contaminants from different families.
• The methodology has been applied to real waters samples.

ABSTRACT

A multiclass wide-scope screening of organic contaminants in natural and waste water has been developed and validated for qualitative purposes, i.e. detection and reliable identification of compounds detected in samples at a certain level of concentration. The screening is based on the use of liquid chromatography coupled to quadrupole-time of flight mass spectrometry (LC–QTOF MS) and has been applied to water samples of different origin and matrix composition (surface water, ground water and effluent urban wastewater). Water samples were spiked with a standard mixture of around 150 organic contaminants from different chemical families (including a number of relevant metabolites/transformation products (TPs), at 0.1 and 1 µg/L concentration levels. After solid-phase extraction with Oasis HLB cartridges, sample extracts were analyzed by LC–QTOF MS and the accurate-mass full-spectrum data were processed for qualitative analysis. The presence of at least two ions (typically the (de)protonated molecule and one fragment ion) accurate-mass measured was used for the reliable identification. The screening detection limit (SDL) and the limit of identification (LOI) were established as the main parameters of the screening method. Nearly all compounds could be detected at the lowest concentration tested, but identification was problematic for some compounds at 0.1 µg/L level, especially in wastewater samples. The screening procedure was finally applied to different water

samples using a home-made database of around 1100 organic contaminants. It allowed the detection and identification of several antibiotics, anti-inflammatory/analgesics drugs and lipid regulators. Cocaine and its metabolite benzoylecgonine were also frequently detected. In addition, triazine herbicides and their TPs, and fungicides like thiabendazol, carbendazim or imazalil, were also identified in some of the samples.

1. INTRODUCTION

Environmental pollution is a matter of increasing concern as the number of contaminants potentially hazardous for plant and animal species, or even for human health, is extraordinary high. During many years, the impact of chemical pollution has been focused almost exclusively on the so-called "priority" pollutants, especially those acutely toxic/carcinogenic pesticides and industrial intermediates displaying persistence in the environment. However, in the last years, the presence of emerging contaminants, such as pharmaceuticals [1-3], illicit drugs [4-6] or personal care products [7-9], is being reported in the environment. These compounds can finally reach the aquatic environment due to the inability of water treatment plants to completely remove some of them [10].

Environmental analytical chemists have to face the investigation of thousands of organic pollutants from different origin and chemical composition that may be present in the aquatic environment. Multiclass/multiresidue methods, commonly based on LC–MS/MS, are applied to monitor most emerging contaminants, as this technique allows the satisfactory identification and quantification of target analytes at low concentrations. However, although quantitative data are highly useful, information on the possible presence of other pollutants that do not form a part of the list is missed in most target methods [11-13].

Full-spectrum acquisition techniques based on the use of high resolution mass spectrometry (HRMS) are increasingly being applied for wide-scope screening purposes in the environmental field [12-15]. Screening a large number of contaminants is highly useful in the environment to have realistic information of the actual composition of samples [16]. HRMS can achieve extended multi-class compounds information as well as provide a broad and realistic knowledge about sample pollution in a rapid, sensitive and selective way [17]. Moreover, this approach facilitates the easy method update or even the retrospective analysis of suspected samples in a subsequent step, a relevant fact as "emerging contaminants" are continuously

appearing (as well as new transformation products discovered), this being another new reason of concern [18].

Typically, a large screening is focussed on qualitative analysis, i.e. the objective is to know whether a given analyte is present or not in the sample at a certain concentration level. Oppositely to other applied fields, like toxicology or anti-doping analysis [19,20], there is a lack of environmental wide-scope screening methods conveniently validated from a qualitative point of view [21,22]. In fact, validation guidelines are commonly focussed on quantitative methods [23], and there is not a widely accepted procedure to be applied in qualitative analysis yet. Recently, SANCO documents have included screening methods based on MS techniques and give some useful indications on their validation in the field of pesticide residue analysis [24]. As the goal of screening is detect and identify the presence of contaminants in the sample, qualitative validation pursues establishing the minimum concentration for each compound that can be detected (screening detection limit, SDL) and identified (limit of identification, LOI) in a reliable way in the different matrices under study. Qualitative validation of screening methods is useful and necessary to assess the applicability and potential limitations of the procedure applied and to improve the knowledge on the analyte concentration limits and on specificity of the method. Due to the lack of information on screening methodologies, some authors have proposed criteria to be achieved when developing screening methodologies with HRMS [22,25].

The objective of this work is to develop and perform the qualitative validation of a screening method for organic pollutants in natural water and wastewater based on the use of ultra-high pressure liquid chromatography coupled to hybrid quadrupole time-of-flight mass spectrometry (UHPLC–(Q)TOF MS). In order to evaluate the methodology, a subset of around 150 organic contaminants from different chemical families has been used. Once validated, the screening method has been applied to the analysis of different water matrices, including ground water, surface water and wastewater, to test its applicability. QTOF under MS^E mode has been applied, which allows the reliable identification of the positive findings in just one run thanks to the simultaneous acquisition of low and high collision energy spectra. Identification criteria when using HRMS with promoted fragmentation in wide-scope screening methods are discussed too.

2. EXPERIMENTAL

2.1. Reagents and chemicals

Reference compounds were purchased from Across Organics (Geel, Belgium), Bayer Hispania (Barcelona, Spain), Fort Dodge Veterinaria (Gerona, Spain), Vetoquinol Industrial (Madrid, Spain), Aventis Pharma (Madrid, Spain), Sigma Aldrich (St. Louis, MO, USA), Cerilliant (Round Rock, TX, USA), Dr. Ehrenstorfer (Augsburg, Germany), Riedel-de Haën (Seelze, Germany), the National Measurement Institute (Pymble, Australia) and Fluka (Buchs, Switzerland). All reference materials presented purity higher than 93%.

Individual stock solutions of each compound (around 500 mg/L, when possible) were prepared in methanol, acetonitrile or acetone. Regarding antibiotics, the addition of 100 µL of 1 M NaOH was necessary for the proper dissolution of the acidic analytes like quinolones. Six intermediate mixed solutions at 1 mg/L were prepared in MeOH. Finally, an intermediate mixed solution containing all compounds at a concentration of 100 µg/L was obtained after mixing the six intermediate solutions and diluting with water. Working solutions were subsequently prepared from the mixed solution by diluting the appropriate volume with HPLC-grade water. All standard solutions (stock, intermediate and working solutions) were stored in amber glass bottles at −20 °C in a freezer.

HPLC-grade water was obtained by purifying demineralised water in a Milli-Q plus system from Millipore (Bedford, MA, USA). HPLC-grade methanol (MeOH), HPLC-grade acetonitrile, residue analysis grade acetone, sodium hydroxide >99% (NaOH), ammonia solution (25%), and formic acid (98–100%) were acquired from Scharlau (Barcelona, Spain). Leucine–enkephalin, used as the lock mass, was purchased from Sigma–Aldrich.

Oasis HLB (60 mg) as well as Oasis MCX (60 mg) cartridges, used for solid-phase extraction, were purchased from Waters (Milford, MA, USA).

2.2. Instrumentation

A Waters Acquity UPLC system (Waters) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (Q-oaTOF Premier, Waters Micromass, Manchester, UK), using an orthogonal Z-spray-ESI interface operating in positive and negative ion mode. The UPLC separation was performed using an Acquity UPLC BEH C18 1.7 µm particle size analytical column 100 × 2.1 mm (Waters) at a flow rate of 300 µL/min. The mobile phases used were A = H2O with 0.01% HCOOH and B = MeOH with 0.01% HCOOH. The percentage of organic modifier (B) was changed linearly as follows: 0 min, 10%; 14 min, 90%; 16 min, 90%; 16.01 min, 10%;

18 min, 10%. Nitrogen was used as the drying gas and nebulizing gas. The gas flow was set at 600 L/h. TOF-MS resolution was approximately 10,000 at full width half maximum (FWHM) in V-mode at m/z 556.2771. MS data were acquired over an m/z range of 50–1000. A capillary voltage of 3.5 kV and 3.0 kV for positive and negative ion modes, respectively, and cone voltage of 25 V were used. Collision gas was argon 99.995% (Praxair, Valencia, Spain). The interface temperature was set to 350 °C and the source temperature to 120 °C. The column temperature was set to 60 °C.

For MSE experiments, two acquisition functions with different collision energies were created: the low energy function (LE), selecting a collision energy of 4 eV, and the high energy (HE) function, with a collision energy ramp ranging from 15 eV to 40 eV in order to promote fragmentation. The LE and HE functions settings were for both a scan time of 0.2 s and an inter-scan delay of 0.05 s. The automated attenuated function was also selected to correct for possible peak saturations (extended mode).

Calibration of mass axis was conducted from m/z 50 to 1000 with a 1:1 mixture of 0.05 M NaOH:5% HCOOH diluted (1:25) with acetonitrile:water (80:20), at a flow rate of 10 µL/min. For automated accurate mass measurement, the lock-spray probe was used, using as lock mass a solution of leucine–enkephalin (2 µg/mL) in acetonitrile:water (50:50) at 0.1% HCOOH pumped at 30 µL/min through the lock-spray needle. A cone voltage of 65 V was selected to obtain adequate signal intensity for this compound ($\square$500 counts). The (de)protonated molecule of leucine–enkephalin, at m/z 556.2771 in positive mode and m/z 554.2615 in negative mode, was used for recalibrating the mass axis and ensuring a robust accurate mass measurement along time. It should be noted that all the exact masses shown in this work have a deviation of 0.55 mDa from the "true" value, as the calculation performed by the MassLynx software uses the mass of hydrogen instead of a proton when calculating $[M+H]^+$ exact mass. However, because this deviation is also applied during mass axis calibration, there is not negative impact on the mass errors presented in this article. MS data were acquired in centroid mode and were processed by the ChromaLynx XS application manager (within MassLynx v 4.1; Waters Corporation).

2.3. Water samples

Three surface water (SW), three ground water (GW) and three effluent wastewater samples (EWW) were collected from different locations of the Castellon province (Spain) and were used for method validation.

The developed procedure was applied to some other water samples to test its applicability. For this purpose, six SW samples were collected at different sites from the Comunidad Valenciana and from Ebro River (Tarragona); six GW samples, corresponding to vulnerable aquifers within areas of intensive agriculture, were also collected from wells in the Comunidad Valenciana; finally, three EWW samples were collected from a wastewater treatment plant in Castellon. All samples were collected in high-density polyethylene bottles and stored in the dark at a temperature below −18 °C until analysis.

2.4. Analytical procedure

To facilitate the procedure overview, Fig. 1 summarizes the overall screening method applied, including both sample treatment as well as LC–MS data processing.
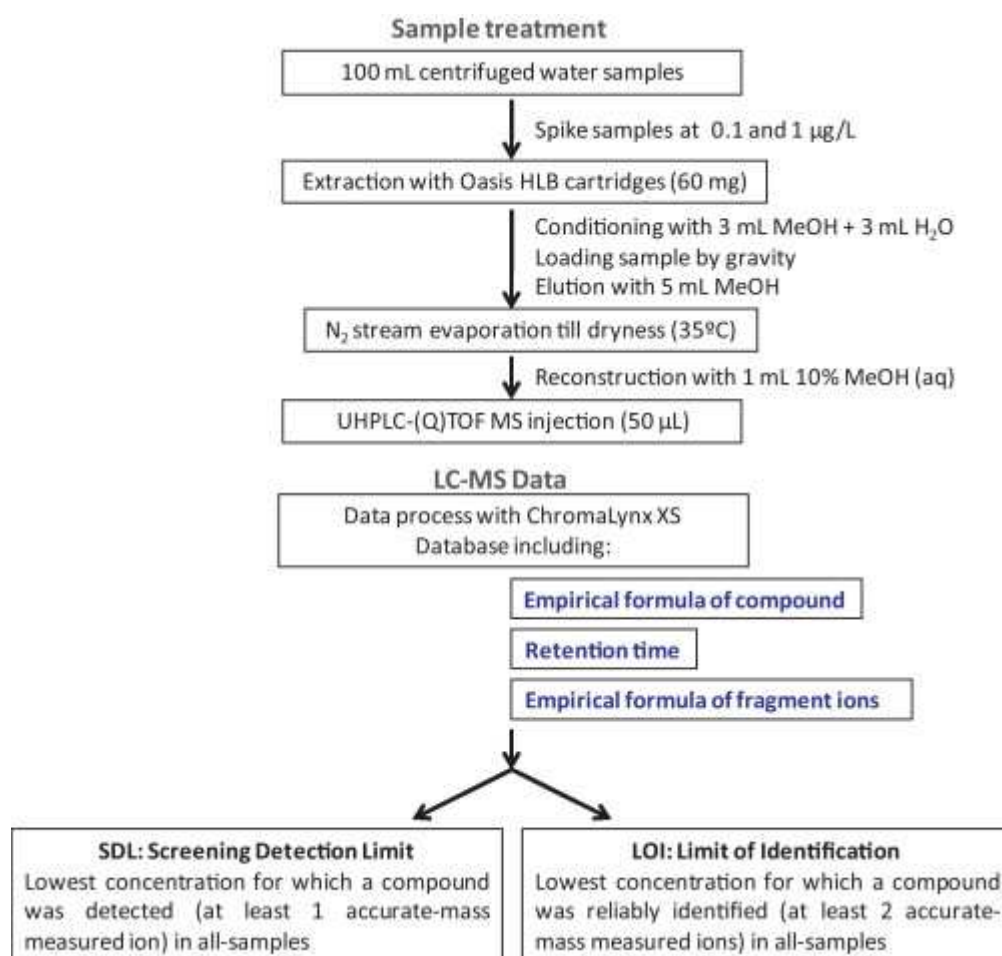

Fig. 1. Overall scheme of the screening method applied.

Water samples were centrifuged at 4500 rpm for 5 min to remove possible solid material. The recommended SPE procedure was based on the use of Oasis HLB cartridges. Briefly, 100 mL of water sample were passed through an Oasis HLB cartridge, previously conditioned by washing and rinsing with 3 mL of MeOH and 3 mL of Milli-Q water. After loading the sample by gravity, the cartridge was air-dried, using vacuum for at least 5 min, and then eluted with 5 mL MeOH. The eluate was evaporated to dryness at 35 °C under a gentle stream of nitrogen, and reconstituted with 1 mL of 10% methanol aqueous solution. Analyses were performed by injecting 50 µL of the final extract into the UPLC–(Q)TOF MS system.

2.5. Qualitative validation

A total of nine water samples (3 SW, 3 GW and 3 EWW) were spiked at two concentration levels each (0.1 and 1 µg/L) and analyzed together with their respective blanks (non-spiked) for qualitative validation. The screening detection limit (SDL) was established as the lowest concentration level tested for which a compound was detected in all the samples, using the most abundant ion (normally, the (de)protonated molecule). The limit of identification (LOI) was established as the lowest concentration tested for which a compound was satisfactorily identified in all spiked samples. The identification criterion was the presence of, at least, two m/z ions in either the LE or HE function, at the expected retention time (2.5% RT deviation tolerance) and a minimum peak width of 5 s, measured at their exact mass (i.e. two peaks in the respective narrow-window extracted ion chromatograms, nw-XIC) with a maximum mass error of 2 mDa.

Selectivity, considered as the ability of the method to discriminate between the analyte and other compounds that could be present in the samples, was tested by determining every analyte in the presence of the rest of compounds included in the screening. It was based on the presence of characteristics m/z ions, measured at accurate mass, for each compound in the LE and HE spectra.

Specificity, considered as the ability of the detector (supported by the selectivity of the extraction, clean-up, derivatization or separation, if applicable) to provide signals that effectively identify the analyte, was checked by analyzing six "blank" natural water samples (SW and GW) and three "blank" wastewater samples. Several of these non-spiked samples contained some of the organic pollutants under study; therefore it was unfeasible to evaluate specificity in these particular cases.

3.   RESULTS AND DISCUSSION

3.1. UHPLC–(Q)TOF MS measurement

In a first step, two SPE procedures based on the use of Oasis MCX [4] and Oasis HLB [3] were evaluated for 120 organic pollutants, including mycotoxins, pharmaceuticals, drugs of abuse, pesticides and relevant metabolites and TPs. Water samples were fortified at 1 µg/L and processed using both type of cartridges.

Using Oasis MCX cartridges, 100 mL of the water sample acidified at pH 2 were loaded onto the cartridges by gravity and, after vacuum-drying 5 min, analytes were eluted with 8 mL of 2% ammonia solution in MeOH. The Oasis HLB procedure has already been described in the experimental section. Recoveries were not taken into account in this study as the goal of the screening was to detect and identify the compounds, not their quantification. The best results were obtained using Oasis HLB cartridges, so it was finally selected as the most adequate. Only the pharmaceutical gabapentin was undetected using HLB but several compounds (e.g. oxilinic, nalidixic and pipedimic acids or moxifloxacin and simvastatin) were not detected when using MCX cartridges.

In a second step, the number of compounds was widened to around 150 to include UV filter agents, preservatives and hormones. In this case, water samples were fortified at both 0.1 and 1 µg/L and processed using only Oasis HLB cartridges. The method was validated for all compounds included in this second study.

As stated in the experimental section, data were acquired under MSE mode, which allows the acquisition of several functions at different collision energies without any m/z pre-selection in the first quadrupole, producing CID fragmentation of all ions generated in the ESI interface. In this work, two functions were acquired, the low collision energy (LE) function, at 4 eV, and the high collision energy (HE), using a collision energy ramp from 15 to 40 eV. In the LE function, mainly (de)protonated and/or sodium adducts are expected although in-source fragments can also been observed in some particular cases, while in the HE function, fragmentation is promoted and fragment ions are more abundant.

Table S1 (Supplementary information) shows the molecular formulae and exact masses for the m/z ions included in the validation for each compound. For up to 58 analytes, it was feasible to use 5 or even more ions, giving an extraordinary power to the identification process.

Accurate-mass full-spectrum acquisition data were processed using an automated processing method (ChromaLynx XS). As this software automatically calculates the (de)protonated exact mass from the molecular formula included in the database, in the case of the confirmatory fragment ions the listed chemical formula differs ±1 hydrogen from the real empirical fragment formulae as a function of the ionization mode employed. Thus, one H was subtracted for positive ESI mode or added for negative ESI mode to finally obtain the correct exact mass of the fragment.

3.2. General aspects of qualitative validation protocol

The aim of a qualitative validation is to ensure the presence of an analyte in a sample at a certain concentration level. As no quantitation is necessary, method recovery, accuracy and precision are not considered. This type of validation has been widely performed in anti-doping analysis [19,20,26], where the use of screening methods is common. However, qualitative validation of screening methods is less frequent in other fields, like pesticide residue analysis [25] and [27] and environment [21]. Recently, SANCO documents on pesticide residue analysis have included some criteria and parameters to be considered in the validation of screening qualitative methods [24].

In this work, method validation has been performed following the same strategy than in our previous paper [21]. As the main purpose of the qualitative screening is to detect and identify positive samples at a given concentration, the method was considered as satisfactorily validated at certain concentration level when the target analyte was detected (SDL) and correctly identified (LOI) in all the nine samples tested, independently on the recovery and precision. Two concentration levels were assayed (0.1 and 1 µg/L) in each water sample.

The identification criteria in screening methods are a matter of present discussion, particularly the compliance of ion intensity ratios. Recently, Mol et al. have made a detailed discussion on this issue when using full-acquisition HRMS and suggested to widen ion ratio tolerance [25]. Furthermore, a stricter retention time tolerance is also proposed. This is in the line of our previous suggestion of revision of the ion ratio tolerances admitted when using GC–TOF MS for

screening, as compliance of ion ratios is rather problematic in environmental samples at low analyte concentrations [21]. In large screening methods, the measurement of the ion ratios becomes non-suitable or much more tedious than in target quantitative methods where a limited number of analytes is included in the scope. In this work, the method performance was evaluated for 150 compounds, but a wider list of approximately 1100 compounds, was tested in analyses of samples. In addition, analyte ions were monitored in different functions (LE and HE), and variations in the fragmentation performance might occur depending on the sample matrix, which notably affects the ion intensity ratios. This makes accomplishment of Q/q intensity ratios troublesome, as they can be modified in the samples more than expected, especially at low analyte concentrations.

As an example, Fig. 2 shows a positive of terbuthylazine in surface water. A Q/q ratio deviation of 39% for the most abundant ion, higher than the allowed 20% tolerance (2002/657/CE) [28], was obtained in this sample. However when the complete HE spectra were compared, a good match was observed supporting the identification of this herbicide. The same behavior was observed in several analyte/matrix combinations. Our results show that Q/q ratio, which are commonly used for identification in LC–MS/MS methods, do not fit-for-purpose when using accurate-mass full-spectrum techniques, such as time-of-flight mass spectrometry. If applied, higher tolerances would be required as recently suggested [25]. Thus, the presence of two accurate-mass measured ions in any of the acquisition functions (low collision energy function and/or high collision energy function) using a narrow mass window of 0.02 Da with mass errors bellow 2 mDa, at expected retention time, has been used in this work as criterion for the reliable identification of the compounds detected in the screening.
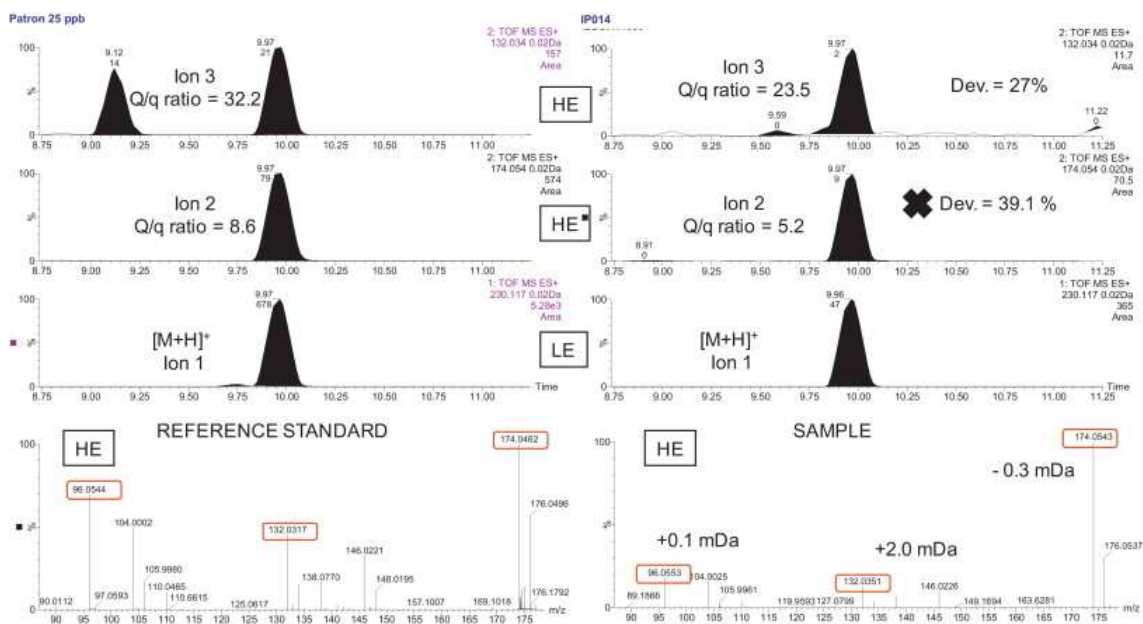
Fig. 2. HE spectra and XICs at 20 mDa mass window for [M+H]+ and main fragment ions (with the corresponding Q/q ratio) of a 25 µg/L reference standard of terbuthylazine (left) and a positive finding of this herbicide in surface water (right).

3.3. Validation results

Three different aqueous matrices were tested in method validation: surface water, ground water and effluent waste water Three samples of each type were spiked at two concentration levels (0.1 and 1 µg/L). It must be emphasized the difficulties to find realistic samples free of any target analyte. Then, those waters previously analyzed and proven to have few positive findings were selected as "blanks" to facilitate the validation process (see Table 1). Although several compounds were found in the "blank" samples, only two contaminants were detected in 3 out of 3 samples of the same type (terbuthylazine in SW and triclosan in SW and EWW). Thus, the method could not be validated for these compounds in this particular kind of samples.

The method was considered fully validated for a given compound, at a certain concentration, when it was detected and correctly identified in all the samples tested (i.e. 9 out of 9). Table 1 summarizes the SDL and LOI obtained for each analyte. It can be seen that the wide majority of compounds (88%) could be detected at the 1 µg/L level while the percentage of detection decreased down to 62% at 0.1 µg/L.

Table 1. Validation results obtained after analysis of nine different spiked samples at different concentration levels and corresponding screening detection limit (SDL) and limit of identification (LOI). Samples for which certain analytes have been detected in the non-spiked blank are rejected for these particular compounds.

Compound	Detected per water type (n=3)						Identified per water type (n=3)						Detected (n=9)		Identified (n=9)		Detected in "blank" samples (n=9)	LOI (μg/L)	SDL (μg/L)
	1 μg/L			0.1 μg/L			1 μg/L			0.1 μg/L			1 μg/L	0.1 μg/L	1 μg/L	0.1 μg/L			
	GW	SW	EW	GW	SW	EW	GW	SW	EW	GW	SW	EW							
11-nor-9-carboxy-Δ9-THC	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
17 beta-estradiol	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	3/9	0/9	0/9	0/9	-	-	1
2-hydroxy-atrazine	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
2-hydroxy-simazine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
2-hydroxy-terbuthylazine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
4-amino-antipyrine	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	-	0.1
Aceclofenac	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Acetamiprid	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Aflatoxin B1	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Aflatoxin B2	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Aflatoxin G1	3/3	3/3	3/3	2/3	1/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	3/9	9/9	0/9	-	1	1
Aflatoxin G2	3/3	3/3	3/3	3/3	1/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	4/9	9/9	0/9	-	1	1
Aflatoxin M1	3/3	3/3	3/3	2/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	3/9	0/9	0/9	-	-	1
Alachlor	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Alprazolam	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Amphetamine	3/3	3/3	3/3	3/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	3/9	9/9	0/9	-	1	1
Atorvastatin	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Atrazine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Azinphos-methyl	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Azithromycin	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Azoxystrobin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Bentazone	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Benzophenone-2 (BP-2)	3/3	3/3	3/3	2/3	1/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	6/9	9/9	0/9	-	1	1
Benzophenone-3 (BP-3)	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	1/3	2/3	1/3	9/9	9/9	9/9	4/9	-	1	0.1
Benzophenone-4 (BP-4)	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Benzoylecgonine	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	8/8	8/8	8/8	8/8	1/9	0.1	0.1
Bezafibrate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1

168

Compound	Detected per water type (n=3)						Identified per water type (n=3)						Detected (n=9)		Identified (n=9)		Detected in "blank" samples (n=9)	LOI (µg/L)	SDL (µg/L)
	1 µg/L			0.1 µg/L			1 µg/L			0.1 µg/L									
	GW	SW	EW	GW	SW	EW	GW	SW	EW	GW	SW	EW	1 µg/L	0.1 µg/L	1 µg/L	0.1 µg/L			
Bromacil	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Buprofezin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Butylparaben	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Carbamazepine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Carbaryl	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Carbendazim	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	8/8	8/8	8/8	8/8	1/9	0.1	0.1
Carbofuran	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Chloramphenicol	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Chlormadinon-17-acetate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Ciprofloxacin	3/3	2/2	3/3	0/3	0/2	0/3	3/3	0/2	0/3	0/3	0/3	0/3	8/8	0/8	3/8	0/8	1/9	-	1
Clarythromycin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Clenbuterol	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Clindamycin	3/3	3/3	1/1	3/3	3/3	1/1	3/3	3/3	1/1	0/3	0/3	0/1	7/7	7/7	7/7	0/7	2/9	1	0.1
Cloxacillin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Cocaethylene	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Cocaine	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	6/6	6/6	6/6	6/6	3/9	0.1	0.1
Cyprodinil	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	1/3	1/2	1/3	8/8	8/8	8/8	3/8	1/9	1	0.1
Deethyl terbuthylazine	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	6/6	6/6	6/6	6/6	3/9	0.1	0.1
Deethyl-atrazine (DEA)	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	2/3	0/3	0/3	9/9	9/9	9/9	2/9	-	1	0.1
Deethyl-terbumeton	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	6/6	6/6	6/6	6/6	3/9	0.1	0.1
Deisopropyl-atrazine (DIA)	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	0/2	0/3	0/3	8/8	8/8	8/8	0/8	1/9	1	0.1
Deoxynivalenol	3/3	3/3	3/3	0/3	0/3	0/3	1/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	1/9	0/9	-	-	1
Diazepam	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Diclofenac	3/3	3/3	3/3	3/3	3/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	6/9	9/9	0/9	-	1	1
Dicloxacillin	3/3	3/3	3/3	1/3	2/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	4/9	9/9	0/9	-	1	1
Diethylstilbestrol	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Dimethoate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Diuron	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Enalapril	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Enrofloxacin	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	0/2	6/6	6/6	6/6	0/6	3/9	1	0.1
Erythromycin (-H2O)	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	2/3	0/3	9/9	9/9	9/9	2/9	-	1	0.1
Erythromycin A	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Estradiol 3-sulfate	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-

Compound	Detected per water type (n=3)						Identified per water type (n=3)						Detected (n=9)		Identified (n=9)		Detected in "blank" samples (n=9)	LOI (μg/L)	SDL (μg/L)
	1 μg/L			0.1 μg/L			1 μg/L			0.1 μg/L			1 μg/L	0.1 μg/L	1 μg/L	0.1 μg/L			
	GW	SW	EW	GW	SW	EW	GW	SW	EW	GW	SW	EW							
Estriol	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Estrone	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Estrone 3-sulfate	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Ethylparaben	2/2	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	0/2	0/3	0/2	7/7	7/7	7/7	0/7	2/9	1	0.1
Fenarimol	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Flumequine	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	0.1	0.1
Fumonisin B1	0/3	2/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	5/9	0/9	0/9	0/9	-	-	-
Fumonisin B2	0/3	2/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	3/9	0/9	0/9	0/9	-	-	-
Furaltadone	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Furazolidone	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	-	0.1
Gemfibrozil	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Hexythiazox	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
HT-2 Toxin	3/3	3/3	3/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	2/9	0/9	0/9	-	-	1
Ibuprofen	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Imazalil	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	0/3	0/2	0/3	8/8	8/8	8/8	0/8	1/9	1	0.1
Imidacloprid	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	3/9	-	0.1	0.1
Iprodione	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Irbesartan	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Isoproturon	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Ketamine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	0.1	0.1
Ketoprofen	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Lincomycin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Lorazepam	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Malathion	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Marbofloxacin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
MCPA	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
MDA	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	9/9	-	1	0.1
MDEA (ecstasy)	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
MDMA	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	0/9	-	0.1	0.1
Medroxyprogesterone-17-acetate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	1/3	0/3	0/3	9/9	9/9	9/9	1/9	-	1	0.1
Megestrol acetate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	1/3	0/3	9/9	9/9	9/9	1/9	-	1	0.1
Melengestrol acetate	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Metalaxyl	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	-	0.1

Compound	Detected per water type (n=3)						Identified per water type (n=3)						Detected (n=9)		Identified (n=9)		Detected in "blank" samples (n=9)	LOI (µg/L)	SDL (µg/L)
	1 µg/L			0.1 µg/L			1 µg/L			0.1 µg/L									
	GW	SW	EW	GW	SW	EW	GW	SW	EW	GW	SW	EW	1 µg/L	0.1 µg/L	1 µg/L	0.1 µg/L			
Methamphetamine	3/3	3/3	3/3	3/3	1/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	4/9	9/9	0/9	-	1	1
Methcathinone	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Methidathion	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	1/3	1/3	0/3	9/9	9/9	9/9	2/9	-	1	0.1
Methiocarb	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Methomyl	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Methylparaben	2/2	2/2	1/1	2/2	2/2	1/1	2/2	2/2	1/1	0/2	0/2	0/1	5/5	5/5	5/5	0/5	4/9	1	0.1
Metolachlor	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Molinate	3/3	3/3	3/3	0/3	0/3	0/3	2/3	2/3	1/3	0/3	0/3	0/3	9/9	0/9	5/9	0/9	-	-	1
Moxifloxacin	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	-	0.1
Nalidixic acid	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Naproxen	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Norbenzoylecgonine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Norcocaine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Norfloxacin	3/3	3/3	3/3	1/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	1/9	0/9	0/9	-	-	1
Ochratoxin Alpha	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Ofloxacin	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	2/2	0/2	0/2	1/2	6/6	6/6	6/6	1/6	3/9	1	0.1
Olanzapine	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Omeprazole	3/3	3/3	3/3	1/3	3/3	3/3	3/3	3/3	3/3	1/3	3/3	0/3	9/9	7/9	9/9	4/9	-	1	1
Oxacillin	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Oxolinic acid	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	0/9	0/9	-	0.1	0.1
Paracetamol	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Paroxetine	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	0/9	0/9	-	-	1
Pefloxacin	3/3	3/3	3/3	3/3	1/3	2/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	7/9	9/9	0/9	-	1	1
Penicillin G	3/3	3/3	3/3	3/3	1/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	7/9	9/9	0/9	-	1	1
Pipemidic acid	3/3	3/3	3/3	2/3	0/3	2/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	0/9	4/9	0/9	-	-	1
Pirimicarb	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Pirimiphos-methyl	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	2/2	7/7	7/7	7/7	7/7	2/9	0.1	0.1
Pravastatin	3/3	3/3	3/3	3/3	0/3	2/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	5/9	9/9	0/9	-	1	1
Propanil	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Propazine	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	0/3	0/3	0/3	9/9	9/9	0/9	0/9	-	-	0.1
Propylparaben	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	2/2	0/3	0/2	0/2	7/7	7/7	7/7	0/7	2/9	1	0.1
Pyridaphenthion	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Pyriproxyfen	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Roxythromycin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1

Compound	Detected per water type (n=3)						Identified per water type (n=3)						Detected (n=9)		Identified (n=9)		Detected in "blank" samples (n=9)	LOI (µg/L)	SDL (µg/L)
	1 µg/L			0.1 µg/L			1 µg/L			0.1 µg/L									
	GW	SW	EW	GW	SW	EW	GW	SW	EW	GW	SW	EW	1 µg/L	0.1 µg/L	1 µg/L	0.1 µg/L			
Salicilic acid	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Sarafloxacin	3/3	2/2	2/2	3/3	0/2	2/2	3/3	2/2	2/2	0/3	0/2	0/2	7/7	5/7	7/7	0/7	2/9	1	1
Simazine	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	6/6	6/6	6/6	6/6	3/9	0.1	0.1
Simvastatin	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	2/2	3/3	2/2	2/2	7/7	7/7	7/7	7/7	2/9	0.1	0.1
Sulfadiazine	3/3	2/3	3/3	0/3	0/3	0/3	2/3	1/3	3/3	0/3	0/3	0/3	7/9	0/9	6/9	0/9	-	-	-
Sulfamethazine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Sulfamethiazole	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/9	0/9	0/9	0/9	-	-	-
Sulfamethoxazole	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
T-2 Toxin	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	9/9	9/9	0/9	-	1	0.1
Terbacil	3/3	3/3	2/2	3/3	0/3	2/2	3/3	3/3	2/2	0/3	0/3	0/2	8/8	5/8	8/8	0/8	1/9	1	1
Terbumeton	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Terbuthylazine*	3/3	0/0	3/3	3/3	0/0	3/3	3/3	0/0	3/3	3/3	0/0	3/3	6/6	6/6	6/6	6/6	3/9	?	?
Terbutryn	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Thiabendazole	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	3/3	1/1	2/2	6/6	6/6	6/6	6/6	3/9	0.1	0.1
Thiobencarb	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Triclosan/Irgasan*	3/3	0/0	0/0	0/3	0/0	0/0	0/3	0/0	0/0	0/3	0/0	0/0	3/3	3/3	3/3	3/3	6/9	-	?
Trimethoprim	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Tylosin A	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1
Valsartan	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	3/3	3/3	2/2	8/8	8/8	8/8	8/8	1/9	0.1	0.1
Venlafaxine	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3	9/9	9/9	9/9	9/9	-	0.1	0.1
Zearalenone	3/3	3/3	3/3	0/3	0/3	0/3	3/3	3/3	3/3	0/3	0/3	0/3	9/9	0/9	9/9	0/9	-	1	1

*Compounds which presence was detected in all the blank samples of a certain type. In this case, SDL and LOI could not be defined.

The reliable identification using two accurate-mass ions was feasible for 74% of compounds at 1 µg/L and for 37% at 0.1 µg/L. The identification of the antibiotic trimethoprim in a EWW sample is shown in Fig. 3. Up to 5 fragment ions were observed for this compound in the sample, even at the lowest spiking level tested. It should be remarked that the majority of unidentified compounds (24% at 1 µg/L level) could not be detected neither, due to the low sensitivity under the analysis conditions or to an insufficient pre-concentration process. For example, using the present methodology some hormones could not be detected (neither validated) at any of the concentrations tested, indicating that specific analysis would be required for these compounds, probably with different mobile phase modifiers.
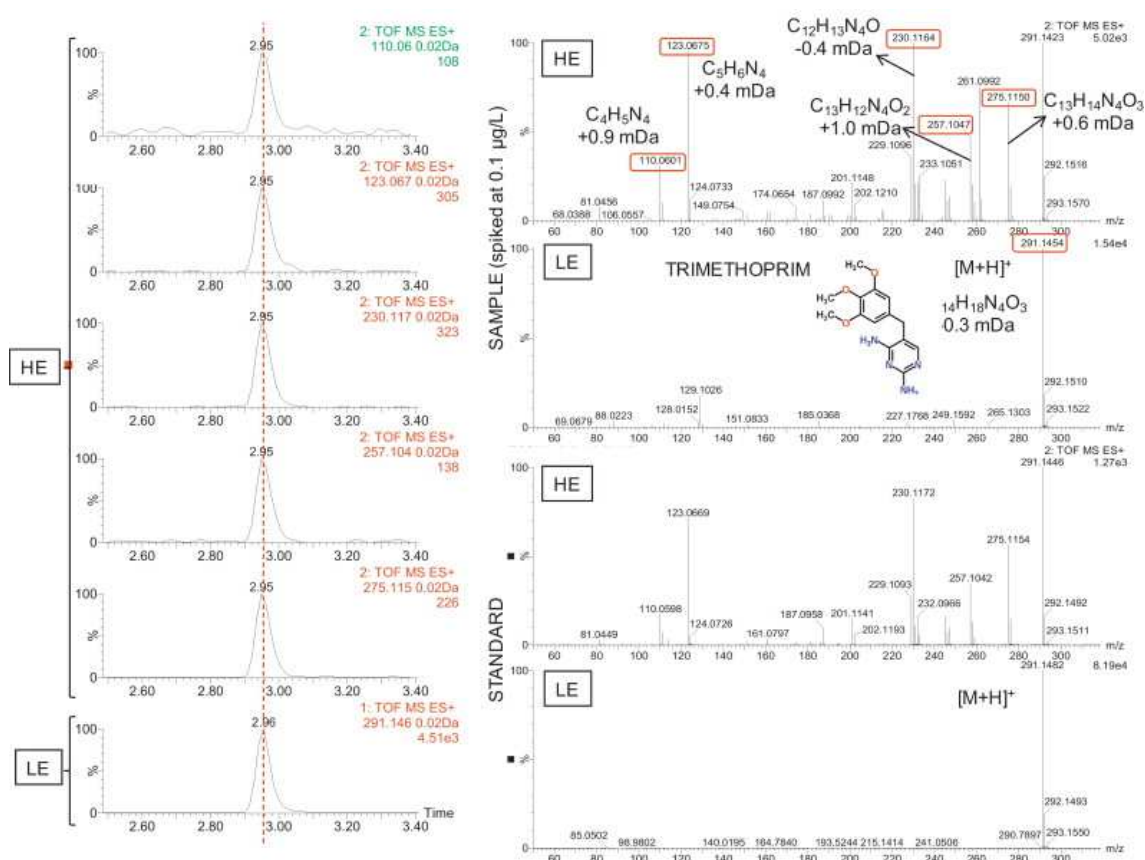
Fig. 3. (left) XICs at 20 mDa mass window for [M+H]+ in LE and main fragments in HE of antibiotic trimethoprim in effluent wastewater; (right) LE and HE mass spectra of the samples and a 50 µg/L reference standard.

In this work, typically the (de)protonated molecule (or adduct) and at least one collision induced dissociation (CID) fragment ion, in either the LE or HE functions, were used. A few compounds (e.g. aceclofenac and some hormones like estriol) presented poor or none fragmentation as a result of the collision energy value applied, which was the result of a compromise for all compounds. For these compounds not enough fragments could be achieved. Therefore, the screening method was limited for detection (only SDL could be set-up), but no identification could be performed. Although some compounds did not present the (de)protonated ion (obtaining adducts or in-source fragments instead), they could be correctly identified when two ions were observed. All the compounds validated in this work, as well as the ions used for detection and/or identification, are shown in Table S1 (Supplementary information).

3.4. Application to routine samples

The developed procedure was applied to the screening of 15 water samples (6 SW, 6 GW, 3 EWW) and the compounds database was widened to around 1100 organic contaminants, including many more pharmaceuticals, pesticides, drugs of abuse, mycotoxins, hormones, UV-filter agents and preservatives, and also some colorants, phenols and surfactants [11]. A notable number of compounds were detected and identified, most of them contaminants that formed a part of the list of compounds subjected to validation, but also some others that were added to the final database used in the screening. Among pharmaceuticals, several antibiotics (ofloxacin, ciprofloxacin and clarithromycin), anti-inflammatory/analgesics drugs (ketoprofen, diclofenac, ibuprofen and salicylic acid) and lipid regulators (gemfibrozil and bezafibrate) were identified using the criterion applied in this work. Regarding drugs of abuse, cocaine and its metabolite benzoylecgonine were the most frequently compounds detected. Triazine herbicides and their transformation products (e.g. simazine, terbumetone, terbuthylazine, deethyl-terbuthylazine and 2-hydroxy-terbuthylazine), and fungicides like thiabendazol, carbendazim and imazalil, were also identified.

When the reference standard was not available in the laboratory, the tentative identification was based on the ions observed (protonated molecule and fragment ions), their compatibility with the chemical structure of the candidate, and by comparison with those ions reported in the literature [11]. This was the case of the diuretic furosemide (Fig. 4), which after tentative identification was finally confirmed after acquisition of the reference standard. Information about fragments was afterwards included to improve the target list in future screenings.
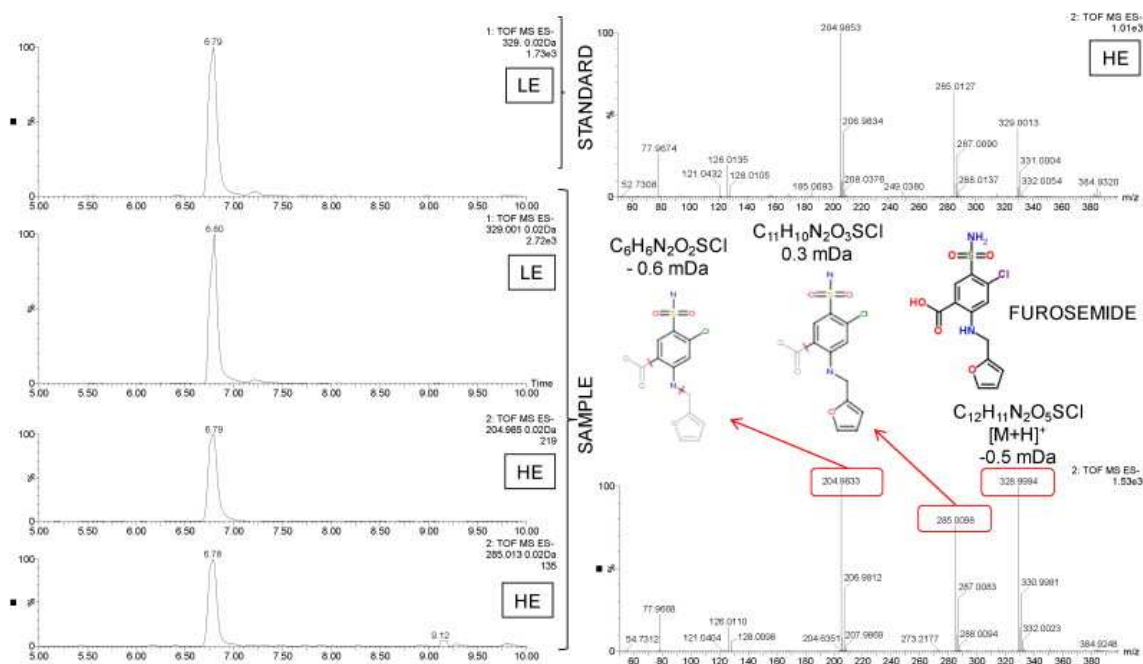
Fig. 4. Positive finding of furosemide in a surface water. XICs at 20 mDa mass window and HE spectra of reference standard and sample, showing proposed fragment ions structures.

As shown in Fig. 5 (terbutryn in a SW sample), some compounds could be identified at concentrations well below the lowest validated level (LOI), which suggest that several contaminants could have been satisfactorily validated at levels below 0.1 μg/L. Following the methodology described in this paper, the use of new-generation QTOF instruments will surely allow the identification of compounds at concentration levels lower than 0.1 μg/L, as suggested by our preliminary data making use of a Xevo G2 QTOF.
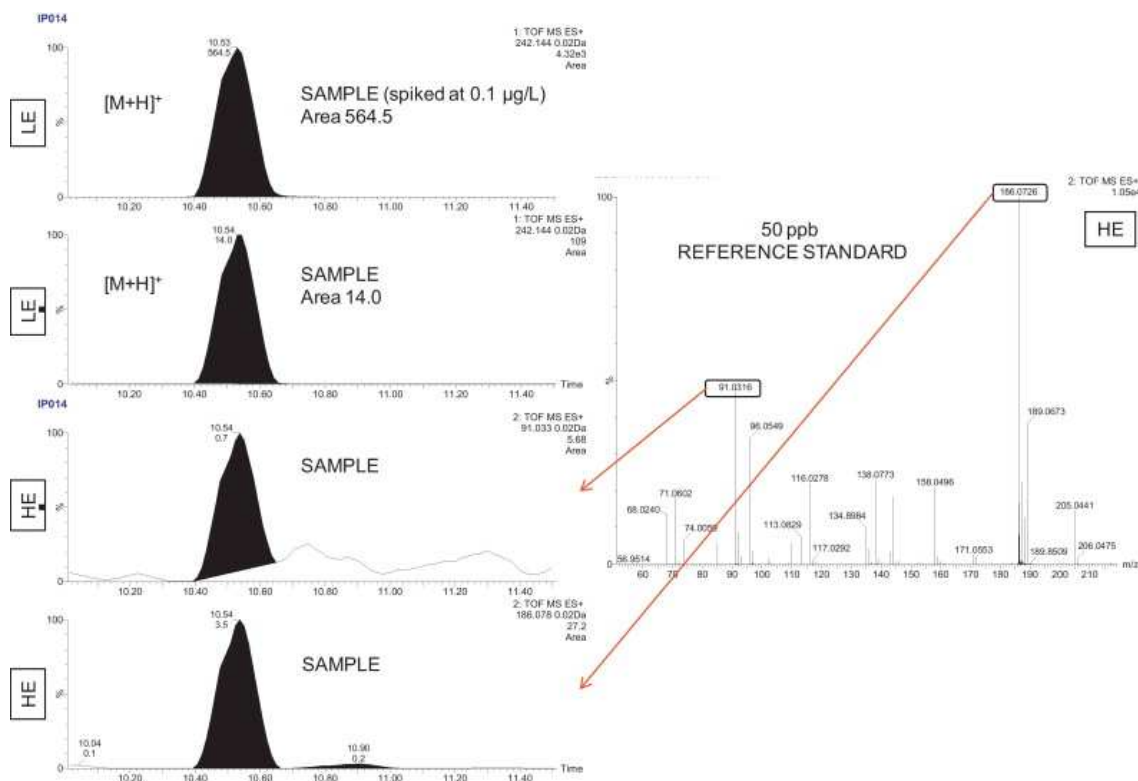
Fig. 5. Identification of the herbicide terbutryn in a surface water at a concentration well below the lowest validated level (0.1 µg/L).

4. CONCLUSIONS

The qualitative validation of a wide-scope screening method including around 150 organic pollutants has been carried out in three types of water samples (groundwater, surface water and effluent waste water). The list of target compounds included many organic contaminants from different families with very distinct physic-chemical properties, such as pesticides and emerging contaminants like pharmaceuticals, drugs of abuse, preservatives or UV filter agents. The SPE pre-concentration step was performed on Oasis HLB cartridges as they provided the best results for the majority of compounds, i.e. the number of detected compound was higher.

Analysis of water extracts by UHPLC–(Q)TOF MS under MSE mode allowed promoting fragmentation of the molecule (without the need of pre-selection of the precursor ion) in a second high-collision energy function. For identification of the compounds, the use of retention

time and the presence of at least two accurate-mass measured ions were applied as a reliable and realistic criterion, as the compliance of ion ratios was problematic using this approach.

Most compounds of the target list were detected at both concentration levels (1 µg/L and 0.1 µg/L) using the most abundant ion (typically the (de)protonated molecule). Unequivocal identification was problematic at the low concentration due to the lower abundance of the fragment ions, but it was feasible for the majority of analytes at the 1 µg/L level. Accurate-mass full-spectrum acquisition in QTOF MS has allowed to easily widen the list of compounds investigated in the analysis of real-world samples. This has facilitated the detection and tentative identification of additional contaminants that were not included in the list of validated compounds.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Fatta-Kassinos, S. Meric, A. Nikolaou, Anal. Bioanal. Chem., 399 (2011), p. 251

[2] S.H. Koo, C.H. Jo, S.K. Shin, S. Myung, Bull. Korean Chem. Soc., 31 (2010), p. 1192

[3] E. Gracia-Lor, J.V. Sancho, F. Hernández, J. Chromatogr. A, 1217 (2010), p. 622

[4] L. Bijlsma, J.V. Sancho, E. Pitarch, M. Ibáñez, F. Hernández, J. Chromatogr. A, 1216 (2009), p. 3078

[5] C. Postigo, M.J. López de Alda, D. Barceló, Environ. Int., 36 (2010), p. 75

[6] J.-D. Berset, R. Brenneisen, C. Mathieu, Chemosphere, 81 (2010), p. 859

[7] M. Pedrouzo, F. Borrull, R.M. Marcé, E. Pocurull, J. Chromatogr. A, 1216 (2009), p. 6994

[8] B. Kasprzyk-Hordern, R.M. Dinsdale, A.J. Guwy, Water Res., 43 (2009), p. 363

[9] D.L. Giokas, A. Salvador, A. Chisvert, TrAC Trends Anal. Chem., 26 (2007), p. 360

[10] M. Gros, M. Petrović, D. Barceló, Environ. Toxicol. Chem., 26 (2007), p. 1553

[11] R. Díaz, M. Ibáñez, J.V. Sancho, F. Hernández, Anal. Methods, 4 (2012), p. 196

[12] F. Hernández, L. Bijlsma, J.V. Sancho, R. Díaz, M. Ibáñez, Anal. Chim. Acta, 684 (2011), p. 87

[13] O.J. Pozo, P. Van Eenoo, K. Deventer, H. Elbardissy, S. Grimalt, J.V. Sancho, F. Hernandez, R. Ventura, F.T. Delbeke, Anal. Chim. Acta, 684 (2011), p. 98

[14] M. Ibáñez, C. Guerrero, J.V. Sancho, F. Hernández, J. Chromatogr. A, 1216 (2009), p. 2529

[15] M.J. Gómez, M.M. Gómez-Ramos, O. Malato, M. Mezcua, A.R. Férnandez-Alba, J. Chromatogr. A, 1217 (2010), p. 7038

[16] A.C. Hogenboom, J.A. van Leerdam, P. de Voogt, J. Chromatogr. A, 1216 (2009), p. 510

[17] M.J. Martínez Bueno, M.M. Ulaszewska, M.J. Gomez, M.D. Hernando, A.R. Fernández-Alba, J. Chromatogr. A, 1256 (2012), p. 80

[18] F. Hernández, M. Ibáñez, E. Gracia-Lor, J.V. Sancho, J. Sep. Sci., 34 (2011), p. 3517

[19] O.J. Pozo, P. Van Eenoo, K. Deventer, F.T. Delbeke, Anal. Bioanal. Chem., 389 (2007), p. 1209

[20] C. Jiménez, R. Ventura, J. Segura, J. Chromatogr. B: Biomed. Anal. Technol. Biomed. Life Sci., 767 (2002), p. 341

[21] T. Portolés, E. Pitarch, F.J. López, F. Hernández, J. Chromatogr. A, 1218 (2011), p. 303

[22] M.W.F. Nielen, M.C. van Engelen, R. Zuiderent, R. Ramaker, Anal. Chim. Acta, 586 (2007), p. 122

[23] S.L.R. Ellison, T. Fearn, TrAC Trends Anal. Chem., 24 (2005), p. 468

[24] European Commission DG-SANCO, Method Validation and Quality Control Procedures for Pesticide Residue Analysis in Food and Feed, No. SANCO/12495/2011, Brussels, 1 October 2012, http://ec.europa.eu/food/plant/protection/pesticides/docs/qualcontrol_en.pdf.

[25] H.G.J. Mol, P. Zomer, M. De Koning, Anal. Bioanal. Chem., 403 (2012), p. 2891

[26] K. Deventer, O.J. Pozo, P. Van Eenoo, F.T. Delbeke, Rapid Commun. Mass Spectrom., 21 (2007), p. 3015

[27] A.G. Frenich, M.J. González-Rodríguez, F.J. Arrebola, J.L. Martínez Vidal, Anal. Chem., 77 (2005), p. 4640

[28] European Union Decision 2002/657/EC, Off. J. Eur. Commun., L221 (12 August 2002) 8.

2.4.1. Discusión de resultados

El tercer artículo de este capítulo representa el último escalón en el desarrollo de una metodología fiable y rápida para la detección e identificaciónde un amplio número de contaminantes orgánicos en diferentes matrices mediante LC-(Q)TOF MS. En este artículo se presenta la metodología aplicada y los resulados obtenidos en la validación cualitativa del screening (detección e identificación de los compuestos). Para ello, se seleccionaron compuestos incluidos en la base de datos, los cuales se tomaron como modelo para la validación, y se añadieron en concentraciones conocidas a distintas muestras de agua. Una vez aplicado el procedimiento de screening a las muestras fortificadas, se pudieron establecer los SDL y los LOI para estos compuestos En el trabajo presentado se hicieron varias aportaciones de interés.

La primera fue corroborar la utilidad y aplicabilidad del método propuesto (en este caso para muestras ambientales de agua), así como el conocer las limitaciones en cuanto a tipo de analitos analizables y las concentraciones que se podían alcanzar. La validación del método permitió comprobar qué compuestos, entre los includidos en la validación, eran detectables/identificables y a qué niveles de concentración. Así, se pudo corroborar que el screening no era aplicable a ciertas familias de analitos como, por ejemplo, algunas hormonas, pero mostró SDLs de 0.1 µg/L (menor nivel de concentración validado) para la mayoría de los compuestos. Además, el análisis de muestras reales mostró que era posible detectar, e incluso identificar, algunos compuestos a niveles inferiores al más bajo validado, pero que no se habían testado

Por otro lado, se evaluaron los criterios propuestos por algunas guías para la correcta identificación para métodos de screening basados en HRMS. Aunque en el campo medioambiental no se dispone de guías para este tipo de validación, se tomaron como referencia algunas guías existentes en otros campos, como por ejemplo la guía SANCO para el análisis de pesticidas en alimentos (SANCO/10684 2009) o la Decision EU 2002/657/EC sobre métodos analíticos para contaminantes y residuos en productos de origen animal (2002/657/EC 2002).

En otros trabajos de nuestro grupo de investigación (Hernández, Ibáñez et al. 2004), ya se discute el sistema de confirmación de la identidad basado en el criterio de IPs en el caso de espectrómetros de masa de alta resolución, ya que desde nuestro punto de vista, dicha clasificación debería establecerse acorde a la exactitud de masa obtenida y no a la resolución del instrumento puesto que, aunque en general ambos parámetros están estrechamente ligados, no son siempre proporcionales. En este artículo, se discutió el criterio usado para la relación de intensidades entre los iones medidos (*ion ratio*) para la identificación de analitos en este tipo de técnicas, y especialmente, para el análisis de muestras mediante el modo MS^E.

Como se explica en el artículo, el cálculo de los ion ratios, resulta tedioso para el screening de contaminantes ya que éste no se centra exclusivamente en la identificación de analitos de los que se dispone de patrón de referencia (y por tanto de tiempo de retención y información de fragmentación) sino que intenta abarcar un elevado número de compuestos, a priori analizables por LC-MS, en base a la esperada formación de ion correspondiente a la molécula (des)protonada. De la mayoría de ellos, no se dispone de patrón de referencia, y por tanto, no se tiene información sobre el *ion ratio* de los iones medidos. Así, la metodología de screening no está dirigida a un número reducido de compuestos y la relación de abundancia de los iones resulta poco aplicable desde un punto de vista práctico. Por otro lado, aunque la fragmentación por MS^E resulta realmente práctica para la obtención de fragmentos para la confirmación de la identidad, como ya se demostró en los anteriores trabajos de este capítulo, esta aproximación puede sufrir interferencias o alteraciones en la relación de la intensidad de iones fragmento en función de la matriz que lleven al incumplimiento de los criterios de tolerancia del ion ratio, y por ende, a reportar falsos negativos. Estos efectos son más acusados a bajas concentraciones. Por ejemplo, en el caso de la terbutilazina, reportado en el artículo, aunque el espectro de HE permita la identificación fiable del compuesto en esta muestra, el criterio de tolerancia en el *ion ratio* hubiera llevado a reportar un falso negativo.

Una de las alternativas propuestas por las guías en estos casos es la adquisición de un ion fragmento adicional, pasando de los dos iones requeridos cuando se calcula el ion ratio a un total de tres iones medidos en masa exacta. Este criterio,

aunque valido, conllevaría un aumento considerable de los LOIs, puesto que se necesitaría observar 2 fragmentos- además de la molécula (des)protonada- para la correcta identificación del compuesto. Según nuestra experiencia, dos iones medidos en masa exacta (tolerancia por debajo de los 2 mDa de error) al tiempo de retención esperado resultaría en un número de falsos positivos mínimo o nulo, probablemente no mejorable con la adquisisción de tercer ion. Además, podría ser incluso mucho más selectivo que el uso de dos transición en equipos de trabajo en m/z nominal, como por ejemplo los analizadores QqQ, cuando no se selecionan iones productos específicos.

Cabe destacar que, aunque los resultados obtenidos para esta validación cualitativa son bastante favorables, el uso de instrumentos más modernos y con características mejoradas, en especial la sensibilidad y exactitud de masa, debería mejorar notablemente los SDLs y LOIs conseguidos. El hecho de que los nuevos equipos QTOF MS presenten mayor exactitud de masa, mejorará la selectividad del screening, pudiendo restringir los errores de masa por debajo de los 2 mDa de error, así como la sensibilidad mediante la aplicación de menores ventanas de masa para la extracción de los cromatrogramas, reduciendo así el ruido y aumentando la relación S/N notablemente.

En cualquier caso, este artículo pone de manifiesto la necesidad de llevar a cabo validaciones cualitativas para los métodos de screening, así como la necesidad de disponer de criterios útiles y prácticos para el análisis cualitativo mediante técnicas HRMS, especialmente en el campo ambiental.

2.5. REFERENCIAS

2002/657/EC, E. U. D. (2002). J. Eur. Commun: 8–36.

Barceló, D. (2003). "Emerging pollutants in water analysis." TrAC Trends in Analytical Chemistry **22**(10): xiv-xvi.

Barceló, D. and M. Petrovic (2007). "Under the analytical spotlight, contaminants emerge: Report on the 2nd EMCO Workshop "Emerging Contaminants in Wastewaters: Monitoring Tools and Treatment Technologies" held in Belgrade, Serbia, 26 and 27 April 2007." TrAC Trends in Analytical Chemistry **26**(7): 647-649.

Benijts, T., R. Dams, W. Lambert and A. De Leenheer (2004). "Countering matrix effects in environmental liquid chromatography-electrospray ionization tandem mass spectrometry water analysis for endocrine disrupting chemicals." J Chromatogr A **1029**(1-2): 153-159.

Bruins, A. P. (1998). "Mechanistic aspects of electrospray ionization." Journal of Chromatography A **794**(1–2): 345-357.

Ellis, J. B. (2006). "Pharmaceutical and personal care products (PPCPs) in urban receiving waters." Environmental Pollution **144**(1): 184-189.

Frye, C. A., E. Bo, G. Calamandrei, L. Calza, F. Dessi-Fulgheri, M. Fernandez, L. Fusani, O. Kah, M. Kajta, Y. Le Page, H. B. Patisaul, A. Venerosi, A. K. Wojtowicz and G. C. Panzica (2012). "Endocrine disrupters: a review of some sources, effects, and mechanisms of actions on behaviour and neuroendocrine systems." J Neuroendocrinol **24**(1): 144-159.

Gracia-Lor, E., M. Martínez, J. V. Sancho, G. Peñuela and F. Hernández (2012). "Multi-class determination of personal care products and pharmaceuticals in environmental and wastewater samples by ultra-high performance liquid-chromatography-tandem mass spectrometry." Talanta **99**: 1011-1023.

Gracia-Lor, E., J. V. Sancho and F. Hernández (2010). "Simultaneous determination of acidic, neutral and basic pharmaceuticals in urban wastewater by ultra high-pressure liquid chromatography-tandem mass spectrometry." Journal of Chromatography A **1217**(5): 622-632.

Hernández, F., M. Ibáñez, E. Gracia-Lor and J. V. Sancho (2011). "Retrospective LC-QTOF-MS analysis searching for pharmaceutical metabolites in urban wastewater." Journal of Separation Science **34**(24): 3517-3526.

Hernández, F., M. Ibáñez, Ó. J. Pozo and J. V. Sancho (2008). "Investigating the presence of pesticide transformation products in water by using liquid chromatography-mass spectrometry with different mass analyzers." Journal of Mass Spectrometry **43**(2): 173-184.

Hernández, F., M. Ibáñez, J. V. Sancho and Ó. J. Pozo (2004). "Comparison of different mass spectrometric techniques combined with liquid chromatography for confirmation of pesticides in environmental water based on the use of identification points." Analytical Chemistry **76**(15): 4349-4357.

Hernández, F., J. V. Sancho, M. Ibáñez, E. Abad, T. Portolés and L. Mattioli (2012). "Current use of high-resolution mass spectrometry in the environmental sciences." Analytical and Bioanalytical Chemistry **403**(5): 1251-1264.

Ibáñez, M., J. V. Sancho, Ó. J. Pozo and F. Hernández (2006). "Use of liquid chromatography quadrupole time-of-flight mass spectrometry in the elucidation of transformation products and metabolites of pesticides. Diazinon as a case study." Analytical and Bioanalytical Chemistry **384**(2): 448-457.

Lehotay, S. J., Y. Sapozhnikova and H. G. J. Mol (2015). "Current issues involving screening and identification of chemical contaminants in foods by mass spectrometry." TrAC - Trends in Analytical Chemistry **69**: 62-75.

Marín, J. M., E. Gracia-Lor, J. V. Sancho, F. J. López and F. Hernández (2009). "Application of ultra-high-pressure liquid chromatography-tandem mass spectrometry to the determination of multi-class pesticides in environmental and wastewater samples: Study of matrix effects." Journal of Chromatography A **1216**(9): 1410-1420.

Petrovic, M., S. Gonzalez and D. Barceló (2003). "Analysis and removal of emerging contaminants in wastewater and drinking water." TrAC Trends in Analytical Chemistry **22**(10): 685-696.

Portolés, T., E. Pitarch, F. J. López and F. Hernández (2011). "Development and validation of a rapid and wide-scope qualitative screening method for detection and identification of organic pollutants in natural water and

wastewater by gas chromatography time-of-flight mass spectrometry." <u>Journal of Chromatography A</u> **1218**(2): 303-315.

Richardson, S. D. (2009). "Water Analysis: Emerging Contaminants and Current Issues." <u>Analytical Chemistry</u> **81**(20): 8654-8654.

SANCO/10684, D. N. (2009). "Method validation and Quality control procedures For Pesticide residues analysis in Food and feed."

Souverain, S., S. Rudaz and J. L. Veuthey (2004). "Matrix effect in LC-ESI-MS and LC-APCI-MS with off-line and on-line extraction procedures." <u>J Chromatogr A</u> **1058**(1-2): 61-66.

CAPÍTULO 3
METABOLÓMICA EN EXHALADOS PULMONARES HUMANOS

CAPÍTULO 3. METABOLÓMICA EN EXHALADOS PULMONARES HUMANOS

3.1. INTRODUCCIÓN

La metabolómica, como ya se explicó en el capítulo de introducción, se ha convertido en una herramienta muy extendida en diversos campos como la biomedicina o la farmacología.

La técnica más empleada ha sido típicamente la NMR, debido a su gran versatilidad, robustez y la información estructural ofrecida (Moco, Vervoort et al. 2007). Así pues, esta técnica presenta una gran variedad de propiedades que la hacen muy ventajosa para este tipo de aplicaciones. La más importante de las características ofrecidas por la técnica NMR es su extraordinario poder para la elucidación de compuestos desconocidos (Jiang and Wang 2014), la cual se ensalza aún más cuando se trabaja con instrumentos de elevada resolución (típicamente por encima de los 600 MHz). Además, este tipo de instrumentos ofrecen también una mayor sensibilidad, ya que en el caso de la NMR, la intensidad de la señal va directamente ligada a la resolución del equipo. Por otro lado, existen diversas aproximaciones aplicables, como por ejemplo ^{1}H, ^{13}C, ^{31}P, ^{19}F NMR, así como la posibilidad de obtención de espectros bidimensionales tipo COSY, STOCSY, etc. que aumentan el rango de metabolitos detectables (convirtiendo a esta técnica en casi universal) y facilitan la identificación de éstos (Sobolev, Mannina et al. 2015). Por tanto, si combinamos la elevadísima información estructural obtenida con la gran cantidad de librerías de espectros disponibles, la NMR permite la elucidación de la gran mayoría de los metabolitos detectados (Wishart 2008).

Por otro lado, la NMR presenta una elevadísima robustez, repetitividad y reproducibilidad, permitiendo analizar las muestras en diferentes días e incluso en

diferentes instrumentos sin que los resultados se vean afectados por ello. Asimismo, otra de las ventajas remarcables que ofrece esta técnica va ligada al hecho de no sufrir pérdidas de sensibilidad durante largas secuencias de análisis. Esto facilita en gran medida el estudio de un elevado número de muestras, como suele ser el caso de las aplicaciones metabolómicas (Moco, Vervoort et al. 2007).

	1999	2000	2001	2002	2003	2004
▭ MS	0	0	1	8	30	56
■ NMR	1	8	9	20	43	63
■ total	3	14	33	77	147	203
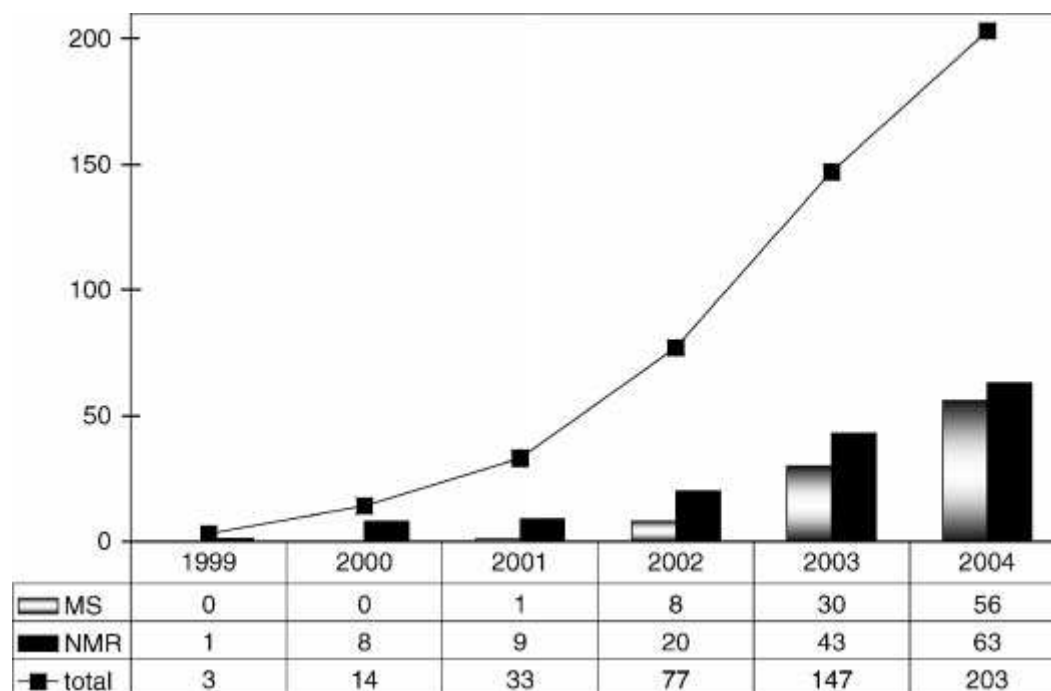
Figura 3.A. Búsquedas bibliográficas en Chemical Abstracts Plus conteniendo las palabras clave metabolomics y metabonomics usando SciFinder Scholar (10 junio 2005). Un total 696 artículos fueron encontrados. El set de datos fue a filtrado usando los parámetros MS o NMR. El diagrama muestra la frecuencia de las publicaciones totales, así como aquellas que mencionan MS y NMR desde 1999 a 2004 (Dettmer, Aronov et al. 2007).

Sin embargo, existen dos principales hándicaps que han hecho irrumpir en los últimos años las técnicas espectrométricas en este tipo de aproximaciones. La primera es el elevado coste de este tipo de instrumentación, no solo con lo que respecta a la compra sino también al mantenimiento y al análisis de muestra, teniendo en cuenta

que se suele trabajar a 4ºK y con disolventes deuterados. La segunda desventaja es la baja sensibilidad ofrecida (incluso por los instrumentos de mayor resolución) en comparación con las concentraciones típicamente detectables mediante GC/LC-MS (Wishart 2008). Como se puede observar en la figura 3.A., entre los años 1999 y 2004, existe un aumento considerable de las publicaciones relacionadas con la metabolómica, pero también una tendencia hacia la paridad de cuanto a las dos principales técnicas usadas. En el año 2014, el número de artículos científicos conteniendo la palabra "metabolomics" se ha incrementado a 2715, mientras que aquellos que además contienen las palabras "MS" o "NMR" son 742 y 408, respectivamente, observándose así una tendencia al alza al uso de MS para estudios metabolómicos.

Así pues, la situación ideal sería aquella que combinara las diferentes posibilidades en cuanto a técnicas se refiere con la finalidad de englobar de forma completa el metaboloma del sistema de estudio, eligiendo las diferentes modalidades de las técnicas NMR para aquellos metabolitos más concentrados y los diferentes acoplamientos de MS para aquellos que se encuentren a baja concentración (Figura 3.B.).

Sin embargo, existen ciertos tipos de matrices que podrían presentar limitaciones de concentración de sus metabolitos en todo el rango. Este es el caso, por ejemplo, de los condensados de exhalados pulmonares (como nos ocupa en este capítulo). Los exhalados pulmonares son una alternativa poco invasiva para el estudio de enfermedades pulmonares y que, por tanto, permite la obtención de un gran número de muestras. Lamentablemente, la baja concentración de los metabolitos en esta matriz dificulta enormemente la obtención de un perfil metabólico fiable y completo. No obstante, la sensibilidad de las diferentes técnicas LC-MS juega un rol importantísimo en estos estudios.
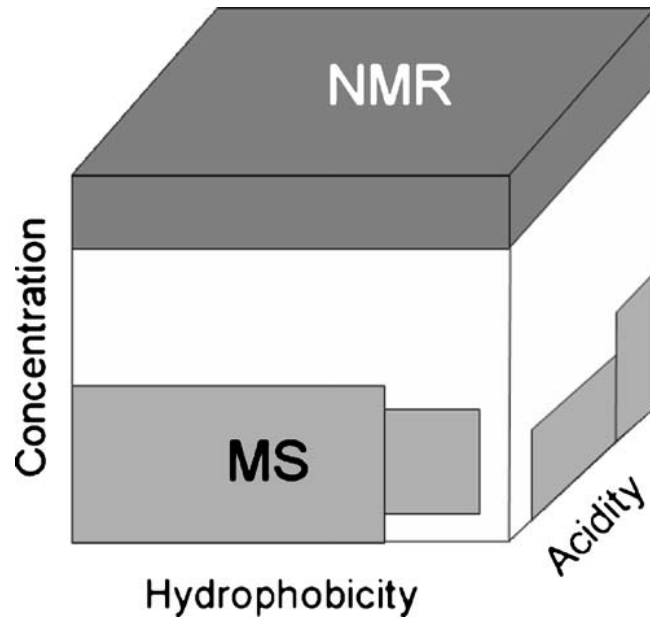
Figura 3.B. Esquema comparativo de las técnicas NMR y MS en función de las diferentes características de los metabolitos.

----------------------- O -----------------------

En este **capítulo,** se explora la posibilidad del uso de condensados de exhalados pulmonares para el estudio de la obstrucción pulmonar de forma no invasiva. Concretamente, se investigan las ventajas y desventajas de las dos principales técnicas instrumentales, comparando su capacidad para afrontar dicho problema.

3.2. Artículo científico 4

"Is NMR-based metabolomic analysis of Exhaled Breath Condensate accurate?"

J.L. Izquierdo-García, G. Peces-Barba, S. Heile, R. Díaz, E. Want and J. Ruiz-Cabello.

European Respiratory Journal, 37 (2011) 468-470.

REFERENCES
1 Rosen SH, Castleman B, Liebow AA. Pulmonary alveolar proteinosis. *N Engl J Med* 1958; 258: 123–1143.
2 Kitamura T, Tanaka N, Watanabe J, *et al.* Idiopathic pulmonary alveolar proteinosis as an autoimmune disease with neutralizing antibody against granulocyte-macrophage colony stimulating factor. *J Exp Med* 1999; 190: 875–880.
3 Kitamura T, Uchida K, Tanaka N, *et al.* Serological diagnosis of idiopathic pulmonary alveolar proteinosis. *Am J Respir Crit Care Med* 2000; 162: 658–662.
4 Inoue Y, Trapnell BC, Tazawa R, *et al.* Characteristics of a large cohort of autoimmune pulmonary alveolar proteinosis patients in Japan. *Am J Respir Crit Care Med* 2008; 177: 752–762.
5 Suzuki T, Sakagami T, Young LR, *et al.* Hereditary pulmonary alveolar proteinosis: pathogenesis, presentation, diagnosis, and therapy. *Am J Respir Crit Care Med* 2010; 182: 1292–1304.

6 Ishii H, Trapnell BC, Tazawa R, *et al.* Comparative study of high-resolution CT findings between autoimmune and secondary pulmonary alveolar proteinosis. *Chest* 2009; 136: 1348–1355.
7 Ministry of Health, Labour and Welfare. Protecting personal identifiable data: ethical guidelines for epidemiological research. Tokyo, Ministry of Health, Labour and Welfare of Japan, 2002; pp. 18–20.
8 Uchiyama M, Nagao T, Hattori A, *et al.* Pulmonary alveolar proteinosis in a patient with Behcet's disease. *Respirology* 2009; 14: 305–308.
9 Lederer DJ, Jelic S, Basner RC, *et al.* Circulating KL-6, a biomarker of lung injury, in obstructive sleep apnoea. *Eur Respir J* 2009; 33: 793–796.
10 Ohnishi T, Yamada G, Shijubo N, *et al.* Secondary pulmonary alveolar proteinosis associated with myelodysplastic syndrome. *Intern Med* 2003; 42: 187–190.
11 Pollack SM, Gutierrez G, Ascensao J. Pulmonary alveolar proteinosis with myeloproliferative syndrome with myelodysplasia: bronchoalveolar lavage reduces white blood cell count. *Am J Hematol* 2006; 81: 634–638.
12 Chung JH, Pipavath SJ, Myerson DH, *et al.* Secondary pulmonary alveolar proteinosis: a confusing and potentially serious complication of hematologic malignancy. *J Thorac imaging* 2009; 24: 115–118.

DOI: 10.1183/09031936.00092910

Is NMR-based metabolomic analysis of exhaled breath condensate accurate?

To the Editors:

The metabolomic analysis of exhaled breath condensate (EBC) is a simple noninvasive approach for the study of respiratory system diseases. Previous studies introduced nuclear magnetic resonance (NMR)-based metabolomics as a method allowing a definite separation between healthy patients and patients with airway disease [1, 2]. In these studies, the influence of external contaminants was also considered. DE LAURENTIIS *et al.* [1] reported that the removal of interfering residual external contaminants was crucial for correct EBC analysis and they proposed a cleaning protocol for the "complete removal of the disinfectant signals" from the reusable parts of the condensers.

In order to verify the influence of the disinfectant signals we have compared the EBC ^{1}H-NMR spectra of a healthy subject obtained after the standard cleaning protocol (disinfected for 15 min using a 1.5% Descogen™ solution and flushed for 15 min with water) recommended by the manufacturer's guidelines and the International Consensus on EBC (fig. 1a) [3]; one that was obtained after the cleaning protocol proposed by DE LAURENTIIS *et al.* [1] (fig. 1b); and one obtained using a device without reusable condenser parts (fig. 1c). EBCs have been collected using an Anacon condenser (Biostec, Valencia, Spain). Sample collection and concentration (>99% water)

were similar to those obtained in the procedure described by DE LAURENTIIS *et al.* [1] and a similar spectrum (fig. 1f) to the spectrum published by DE LAURENTIIS *et al.* [1] was obtained (fig. 1f, insert). Figure 1 shows that although many of the disinfectant signals present in the EBC spectrum have been removed by the DE LAURENTIIS *et al.* [1] cleaning protocol, there are some resistant peaks that are not observed in the spectrum in figure 1c. Furthermore, an EBC ^{1}H-NMR spectrum of a healthy subject acquired using the device without reusable parts (fig. 1d) was compared with another control test (fig. 1e), collected in the same time scale using the condensation of room air pushed into the system by a continuous air spam device using flow and volumes similar to those typical of normal human respiratory function. In this case, it showed similar blank spectra with only two signals (lactate and an unidentified signal).

In our opinion, this comparison indicates that previous studies were not able to remove all disinfectant signals *via* the cleaning protocols. In addition, when the use of disinfectants was circumvented using a non-reusable device, we obtained blank spectra without metabolic information. We are persuaded to conclude that within the normal time periods used for EBC collection (10 min), NMR-based metabolomics does not have

Is NMR-based metabolomic analysis of Exhaled Breath Condensate accurate?

J.L. Izquierdo-García[a], G. Peces-Barba[b], S. Heile[b], R. Díaz[c], E. Want[d] and J. Ruiz-Cabello[a].

[a]*CIBERES, CIBER Enfermedades Respiratorias, Instituto de Estudios Biofuncionales, UCM, Madrid, Spain*
[b]*CIBERES, CIBER Enfermedades Respiratorias, Fundación Jimenez Diaz-CAPIO, Madrid,Spain*
[c]*Research Institute for Pesticides and Water, Universitat Jaume I, Castelló, Spain*
[d]*Department of Biomolecular Medicine, Imperial College London, London, UK.*

To the Editors:

The metabolomic analysis of exhaled breath condensate (EBC) is a simple noninvasive approach for the study of respiratory system diseases. Previous studies introduced nuclear magnetic resonance (NMR)-based metabolomics as a method allowing a definite separation between healthy patients and patients with airway disease [1, 2]. In these studies, the influence of external contaminants was also considered. DE LAURENTIIS et al. [1] reported that the removal of interfering residual external contaminants was crucial for correct EBC analysis and they proposed a cleaning protocol for the ''complete removal of the disinfectant signals'' from the reusable parts of the condensers.

In order to verify the influence of the disinfectant signals we have compared the EBC ^{1}H-NMR spectra of a healthy subject obtained after the standard cleaning protocol (disinfected for 15 min using a 1.5% Descogen^{TM} solution and flushed for 15 min with water) recommended by the manufacturer's guidelines and the International Consensus on EBC (fig. 1a) [3]; one that was obtained after the cleaning protocol proposed by DE LAURENTIIS et al. [1] (fig. 1b); and one obtained using a device without reusable condenser parts (fig. 1c). EBCs have been collected using an Anacon condenser (Biostec, Valencia, Spain). Sample collection and concentration (>99% water) were similar to those obtained in the procedure described by DE LAURENTIIS et al. [1] and a similar spectrum (fig. 1f) to the spectrum published by DE LAURENTIIS et al. [1] was obtained (fig. 1f, insert). Figure 1 shows that although many of the disinfectant signals present in

the EBC spectrum have been removed by the DE LAURENTIIS et al. [1] cleaning protocol, there are some resistant peaks that are not observed in the spectrum in figure 1c. Furthermore, an EBC ^{1}H-NMR spectrum of a healthy subject acquired using the device without reusable parts (fig. 1d) was compared with another control test (fig. 1e), collected in the same time scale using the condensation of room air pushed into the system by a continuous air spam device using flow and volumes similar to those typical of normal human respiratory function. In this case, it showed similar blank spectra with only two signals (lactate and an unidentified signal).

In our opinion, this comparison indicates that previous studies were not able to remove all disinfectant signals via the cleaning protocols. In addition, when the use of disinfectants was circumvented using a non-reusable device, we obtained blank spectra without metabolic information. We are persuaded to conclude that within the normal time periods used for EBC collection (10 min), NMR-based metabolomics does not have the degree of sensitivity needed to detect the lung metabolic fingerprint present in these samples. For this reason, we propose a mass spectrometry (MS) approach for the metabolomic analysis of EBC.

MS-based metabolomics, here using ultra-performance liquid chromatography coupled to a time-of-flight mass spectrometer (UPLC-TOF-MS) requires similar sample preparation with a high degree of sensitivity. Through this technique, we can detect the metabolic fingerprint of the EBC not detected by ^{1}H-NMR spectroscopy. Figure 2 shows the UPLC-TOF mass matrix of an EBC from a healthy subject acquired using the non-reusable device. With the aim of validating the MS-based metabolomics as a discriminatory method between healthy patients and patients with airway disease, preliminary data of a chronic obstructive pulmonary disease (COPD) metabolomic study are presented. A total of 28 EBC samples, 18 COPD and 10 healthy subjects were collected using the non-reusable device and the Anacon condenser. Lyophilised samples were diluted with 200 μL of water and transferred into 96-well plates and maintained at 4°C prior to injection of 5 μL onto the column. Chromatographic separations were performed on a Acquity UPLC system (Waters Corp, Milford, MA, USA) using a 10 min gradient with mobile phase A of 0.1% formic acid (FA) in water and mobile phase B, 0.1% FA in acetonitrile. Electrospray (ESI) mass spectrometry experiments were performed on a Micromass LCT Premier (Waters) operated in positive ionisation mode. The raw data were processed using the open-source freeware XCMS (Scripps Center for Mass Spectrometry, La Jolla, CA, USA), which is able to detect and align metabolite peaks. Spectral differences between COPD and healthy subjects were verified by principal components analysis and partial least squares analysis as shown in figure 3.
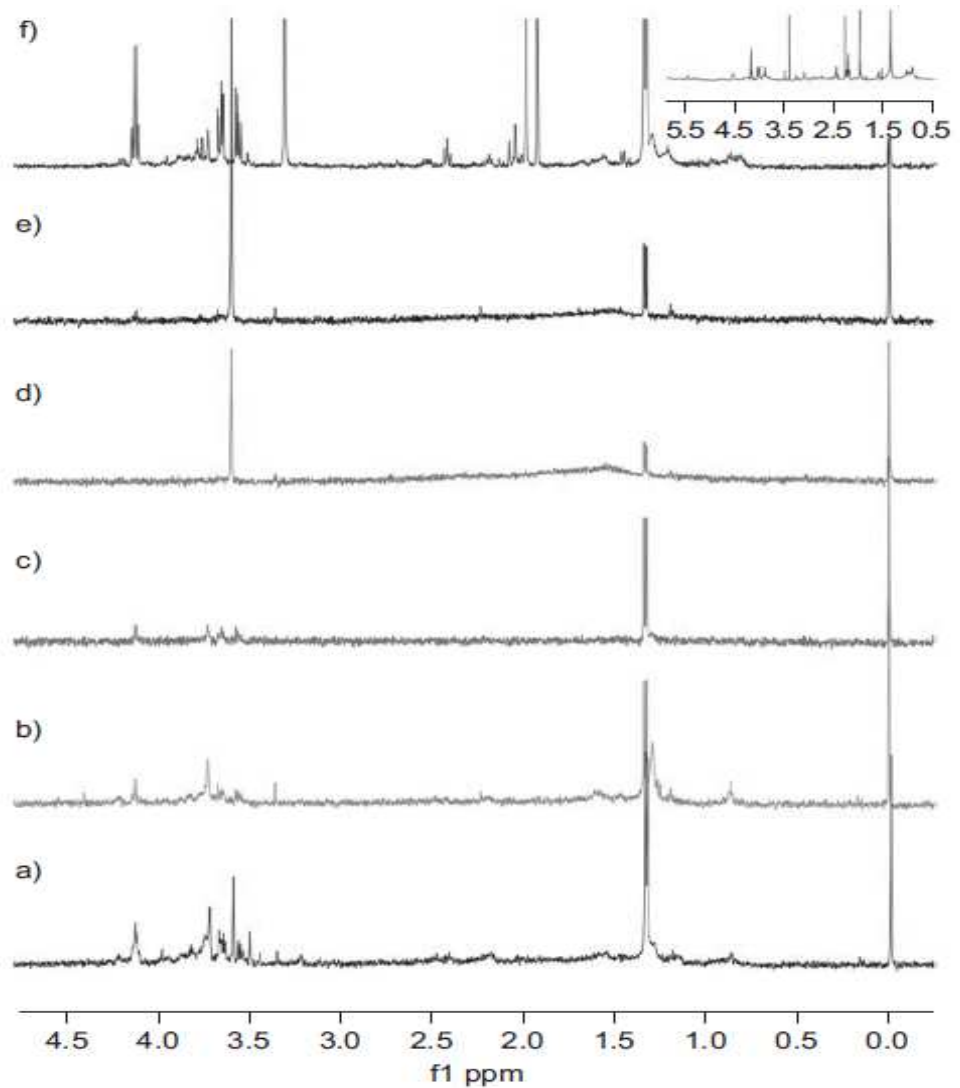
Figure 1. [1]H-nuclear magnetic resonance (NMR) spectra of exhaled breath condensate (EBC) from a healthy subject after a) the standard cleaning protocol recommended by manufacturer's guidelines and the International Consensus on EBC; b) the cleaning protocol proposed by DE LAURENTIIS et al. [1]; c) and d) using a device without reusable condenser parts; and e) ^{1}H-NMR spectra of the condensation of the room air pushed into the system by a continuous air spam device. EBC spectra a), b) and c) were collected from the same healthy subject and spectra c) and d) were collected in the same time scale. f) ^{1}H-NMR spectrum of EBC with similar sample collection and disinfection to ^{1}H-NMR spectrum of EBC published by DE LAURENTIIS et al. [1] (insert; reproduced from [1] with permission from the publisher). f1: first frequency dimension.
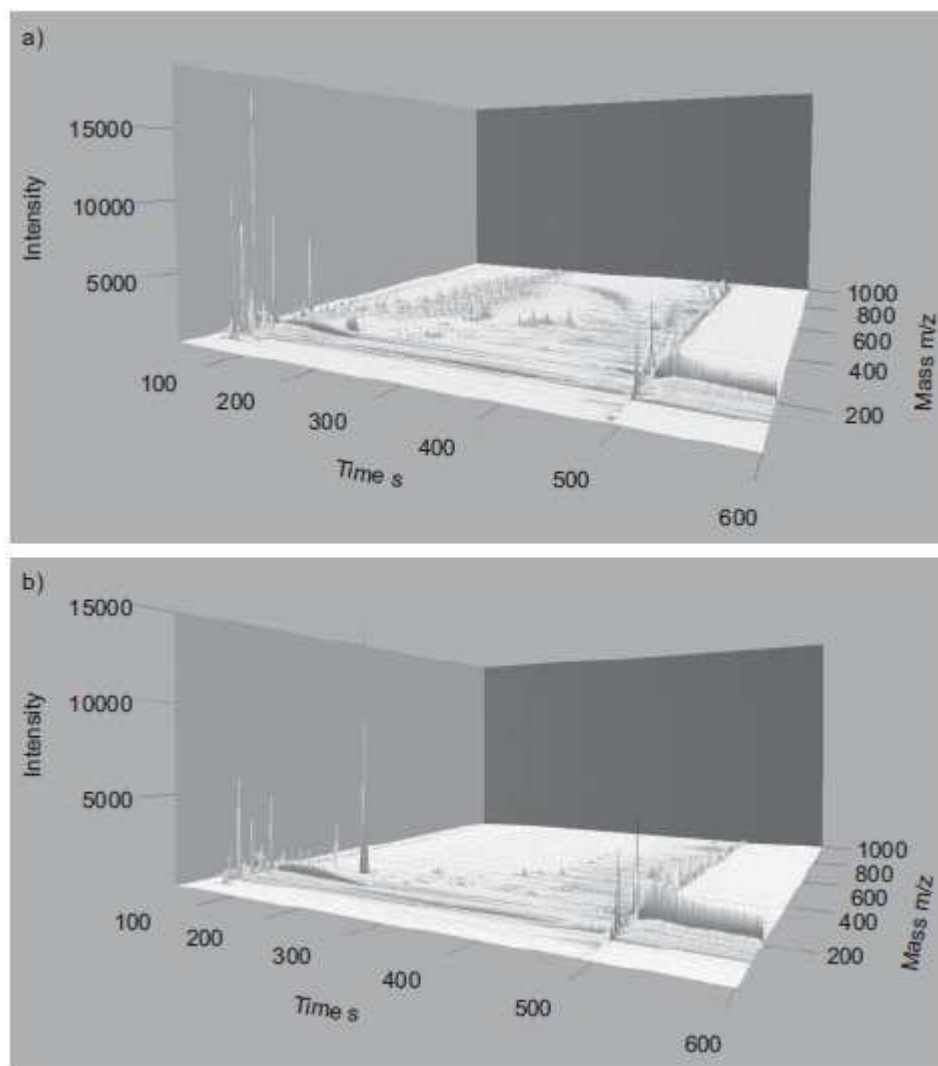
Figure 2. An ultra-performance liquid chromatography coupled to a time-offlight mass spectrometer three-dimensional mass matrix of an exhaled breath condensate from a) a healthy subject and b) a subject with chronic obstructive pulmonary disease, acquired using the non-reusable device. m/z: mass to charge.
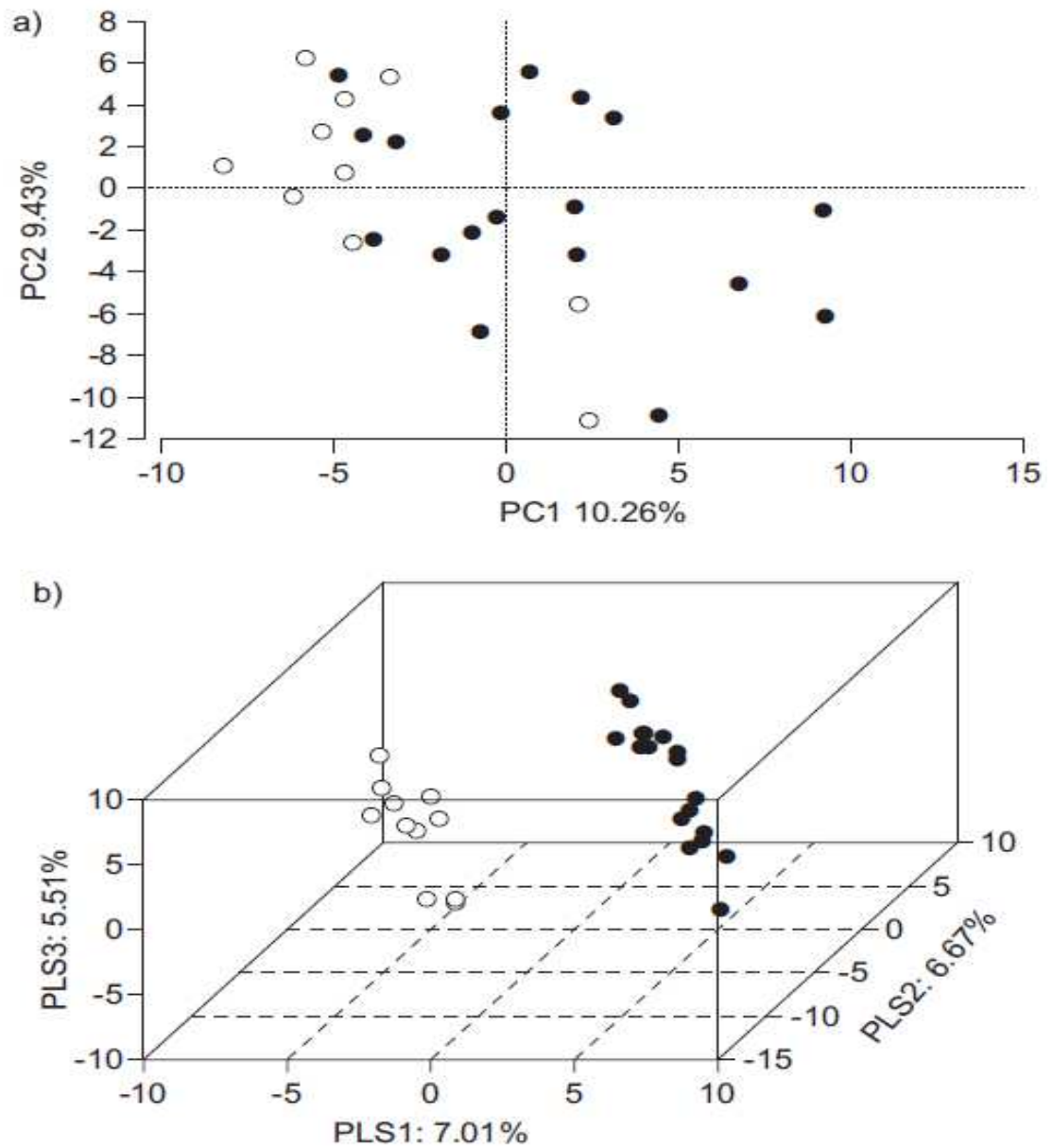
Figure 3. Multivariate analysis using ultra-performance liquid chromatography coupled to a time-of-flight mass spectrometer spectra of exhaled breath condensate from healthy and chronic obstructive pulmonary disease (COPD) subjects. a) Principal components analysis (PCA) score plot shows a separation between groups, which is clearly confirmed by b) partial least squares (PLS) analysis. ○ control; ● COPD. PC1 and 2 are the first two principal components; PLS1 and 2 are the first two partial least squares.

In conclusion, the present study demonstrates that the cleaning protocols of the reusable condenser parts produce an artificial metabolic fingerprint not related to the endogenous metabolic pathway of the lungs. In addition, although NMR-based metabolomics have been successfully applied in the study of numerous biosamples (e.g. urine [4–6] and tissues [7]), this technique cannot be used to analyse EBC samples because the spectroscopic techniques do not have the sensitivity required to observe the endogenous metabolites presented in the EBC. Finally, we propose UPLC-MS and the use of non-reusable devices as a standard metabolomic approach in the analysis of EBC.

REFERENCES

[1] de Laurentiis G, Paris D, Melck D, et al. Metabonomic analysis of exhaled breath condensate in adults by nuclear magnetic resonance spectroscopy. *Eur Respir J* 2008; 32: 1175–1183.

[2] Carraro S, Rezzi S, Reniero F, et al. Metabolomics applied to exhaled breath condensate in childhood asthma. *Am J Respir Crit Care Med* 2007; 175: 986–990.

[3] Horvath I, Hunt J, Barnes PJ, et al. Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J* 2005; 26: 523–548.

[4] Slupsky CM, Rankin KN, Fu H, et al. Pneumococcal pneumonia: potential for diagnosis through a urinary metabolic profile. *J Proteome Res* 2009; 8: 5550–5558.

[5] Saude EJ, Obiefuna IP, Somorjai RL, et al. Metabolomic biomarkers in a model of asthma exacerbation: urine nuclear magnetic resonance. *Am J Respir Crit Care Med* 2009; 179: 25–34.

[6] McClay JL, Adkins DE, Isern NG, et al. [1]H nuclear magnetic resonance metabolomics analysis identifies novel urinary biomarkers for lung function. *J Proteome Res* 2010; 9: 3083–3090.

[7] Izquierdo-García JL, del Puerto-Nevado L, Peces-Barba G, et al. A metabonomic approach to evaluate COPD in a model of cigarette smoke exposure in mice. *Metabolomics* 2010; 6: 564–573.

3.2.1. Discusión de resultados

Este trabajo refleja la importancia del uso de la técnica adecuada para el análisis de las diferentes muestras y compara las dos principales aproximaciones utilizadas para afrontar estudios metabolómicos. Típicamente se ha preferido el análisis mediante NMR debido a las numerosas ventajas que esta técnica presenta y que ya se comentaron en el apartado de 3.1 de introducción del capítulo. Sin embargo, la mayor sensibilidad de las técnicas espectrométricas permite que éstas deban ser seriamente consideradas como una alternativa a la NMR, no solo en casos puntuales como el de muestras de exhalados pulmonares, cuya baja concentración de los analitos requiere del análisis por LC-MS para su determinación (como ha sido puesto en evidencia en este artículo científico), sino en general para todo tipo de estudios, evitando la posible pérdida de información de los metabolitos a menores concentraciones.

Como se comenta brevemente en el artículo, el acoplamiento LC-MS requiere de protocolos para el tratamiento de muestra similares en dificultad y tiempo (o incluso más sencillos) a aquellos aplicados en técnicas espectroscópicas. En este trabajo, las muestras liofilizadas procedentes del condensado de exhalado pulmonar fueron simplemente disueltas en 200 µL de agua y homogenizadas durante 5 minutos mediante ultrasonidos. La simplicidad de los protocolos es esencial para este tipo de técnicas por varias razones: el número de muestras con las que se suele trabajar es muy elevado, se evitan así errores durante el tratamiento de la muestra, además de minimizar las pérdidas en la recuperación de los analitos. Así pues, protocolos extensos generalmente conllevan la pérdida de ciertos analitos por degradación, evaporación o simplemente incompatibilidad con el método de extracción. Como se concluye de este trabajo, la preparación de la muestra consiste en una simple dilución en agua, para posteriormente inyectar las muestras en el sistema LC-MS.

Después de la disolución de la muestra con agua, se dividió la muestra en dos alícuotas: una de 140 µL para trasferir al vial que sería inyectado y otra 40 µL de cada muestra que fueron mezclados con la finalidad de obtener un QC. Por tanto, en este caso, el QC consistiría en una muestra promedio de todas las muestras a analizar. La

finalidad de esta muestra control es la de evaluar posibles errores sistemáticos durante la inyección y el procesamiento de datos, así como comprobar la validez del modelo estadístico generado. Probablemente la principal desventaja del acoplamiento cromatografia-espectrometria de masas, y en especial la cromatografía liquida, frente al NMR es, como se mencionó brevemente en la introducción de este capítulo, la reproducibilidad y repetitividad en las medidas. Durante el análisis de las muestras, los instrumentos suelen sufrir variaciones de sensibilidad y tiempo de retención debido sobre todo a las largas secuencias de inyección requeridas para evitar diferencias en el análisis en días posteriores. Así pues, esta menor repetitividad debe ser evitada mediante el análisis aleatorio y la inyección inicial de al menos 10 veces el QC. Así, se favorece la estabilización de la columna cromatográfica y la fuente de ionización. Además, inyecciones periódicas durante la secuencia de estas muestras QC proporcionan una eficaz herramienta de evaluación de la calidad de los resultados obtenidos, debiendo estar éstas perfectamente agrupadas y centradas (por tratarse de una muestra promedio) en el análisis multivariante. Para minimizar el efecto de la pérdida paulatina de sensibilidad de la fuente electrospray, las muestras fueron inyectadas de forma aleatoria. Se evaluó la opción de normalizar las muestras, pero debido a la casi inexistencia de matriz en los exhalados, la normalización no fue necesaria en base a los resultados de los QC.

Puesto que la mezcla se realiza de forma posterior a la extracción, en ningún caso corregiría errores en el proceso de tratamiento de muestra. Para evitar estos errores, todo el proceso se lleva a cabo de manera aleatoria, al igual que la inyección de las muestras.

Aunque no se comentó en el texto, una de las claves para poder analizar matrices tan diluidas en concentración de metabolitos como los exhalados pulmonares es el uso del acoplamiento LC-MS bajo condiciones óptimas. Para ello, resulta altamente favorable el uso de fases móviles y modificadores del mayor grado de pureza posible, así como una intensa limpieza del la fuente de ionización. Esta es una premisa válida para cualquier análisis non-target, pero crucial en este tipo de matrices pues resulta imprescindible eliminar cualquier compuesto exógeno que podría llevarnos fácilmente a obtener falsos biomarcadores además de complicar en

gran medida el proceso de extracción de picos (peak picking) al enmascarar fácilmente cualquier compuesto endógeno. Este hecho es también aplicable al proceso de toma de muestra (como se subraya en el texto) y su posterior tratamiento.

De este articulo científico se concluye la incapacidad de las técnicas NMR para los estudios metabolómicos de condensados de exhalados pulmonares. Por otro lado, aunque se trate de un estudio preliminar por el número de muestras analizadas y la falta de validación del estudio, se demuestra la validez del acoplamiento LC-MS para el análisis de exhalados pulmonares en el estudio de enfermedades respiratorias, dejando una puerta abierta al uso de este tipo de muestras no-invasivas para el paciente.

3.5. REFERENCIAS

Dettmer, K., P. A. Aronov and B. D. Hammock (2007). "Mass spectrometry-based metabolomics." <u>Mass Spectrometry Reviews</u> **26**(1): 51-78.

Jiang, C. Y. and Y. H. Wang (2014). "Quantitative metabolomics based on NMR." <u>Yaoxue Xuebao</u> **49**(7): 949-955.

Moco, S., J. Vervoort, S. Moco, R. J. Bino, R. C. H. De Vos and R. Bino (2007). "Metabolomics technologies and metabolite identification." <u>TrAC Trends in Analytical Chemistry</u> **26**(9): 855-866.

Sobolev, A. P., L. Mannina, N. Proietti, S. Carradori, M. Daglia, A. M. Giusti, R. Antiochia and D. Capitani (2015). "Untargeted NMR-based methodology in the study of fruit metabolites." <u>Molecules</u> **20**(3): 4088-4108.

Wishart, D. S. (2008). "Quantitative metabolomics using NMR." <u>TrAC - Trends in Analytical Chemistry</u> **27**(3): 228-237.

CAPÍTULO 4
APROXIMACIONES METABOLÓMICAS PARA LA AUTENTIFICACIÓN DE LA DENOMINACIÓN DE ORIGEN DE ALIMENTOS

CAPÍTULO 4. APROXIMACIONES METABOLÓMICAS PARA LA AUTENTIFICACIÓN DE LA DENOMINACIÓN DE ORIGEN DE ALIMENTOS

4.1. INTRODUCCIÓN

La metabolómica, como ya se ha explicado anteriormente, es uno de los campos del emergente grupo de las "-ómicas" centrado en la caracterización masiva de metabolitos en matrices biológicas. Típicamente utilizada en ramas de la biomedicina y la farmacología, también emerge en el área de la ciencia de los alimentos y la nutrición en investigaciones sobre la composición de alimentos y su calidad/autenticidad (German, Hammock et al. 2005).

Este increíble avance ha sido propiciado no sólo gracias a los recientes avances tecnológicos en la instrumentación analítica (especialmente aquella que respecta al análisis de metabolitos pequeños), sino también en gran medida a los desarrollos en los campos del tratamiento de dicho flujo masivo de datos, con la aparición de herramientas informáticas capaces de extraer la información comprendida en los resultados instrumentales, todo ello en una forma muy eficaz y en tiempos extremadamente cortos (Dunn, Bailey et al. 2005, Wishart 2008).

Igualmente importante para el desarrollo de la técnicas metabolómicas ha sido la compilación de bases de datos electrónicas, las cuales contienen información descriptiva y espectral de los constituyentes encontrados en una gran variedad de metabolomas y matrices (Kopka, Schauer et al. 2005, Cotter, Maer et al. 2006, Moco, Tseng et al. 2006, Wishart, Tzur et al. 2007)

Como se comentó en el capítulo de introducción, existen dos aproximaciones principales en el campo de la metabolómica: untargeted (también conocida como metabolomic fingerprint o chemometric approach) y targeted (metabolite profilling o

quantitative approach)(Wishart 2008). En esta tesis se ha hecho uso exclusivamente de la aplicación untargeted, pues el principal objetivo fue encontrar marcadores que permitirían asegurar la autenticidad de determinadas muestras de alimentos de interés.

En el capítulo de introducción se ha hablado extensamente de la importancia del correcto diseño del muestreo, la toma y conservación de la muestra, del tratamiento de la misma y de la técnica analítica utilizada. Sin embargo, el tratamiento de datos es también un paso crítico para el éxito de cualquier proyecto metabolómico. Así pues, a continuación se comentarán algunos de los puntos claves a tener en cuenta.

4.1.1 Pre-procesamiento de datos

Los complejos y multidimensionales conjuntos de datos en metabolómica necesitan un minucioso tratamiento, ya que los pasos de pre-procesamiento de datos empleados tienen un gran impacto tanto en la extensión y calidad de la identificación de metabolitos, así como en las capacidades de cuantificación posteriores y la interpretación biológica. Los factores clave en el pre-procesamiento de datos son: i) la calidad de los datos en bruto generados, ii) la elección del software de tratamiento previo, iii) la optimización de los parámetros del software, y iv) el método de normalización de datos.

Se entiende como pre-procesamiento de datos al proceso de extracción de datos a partir de los resultados del análisis. Centrándonos en el acoplamiento cromatografía-MS, dicho procesamiento de datos incluye las etapas de peak picking, alineamiento del tiempo de retención, integración/cuantificación y normalización de los datos (si se intuye necesario). En cuanto al software para dicho pre-procesamiento de datos, existen diferentes opciones disponibles, tanto comerciales como gratuitas. Obviamente, los algoritmos empleados en las diferentes etapas (detección de picos, alineamiento de tiempos de retención,...) pueden divergir. La Tabla 4.A resume algunas de las principales opciones.

Tabla 4.A. Software comercial (A) y libre (B) para el (pre)procesamiento de datos metabolomic de MS (Want 2011).

A)

Software	Manufacturer	Capabilities
Bluefuse	BlueGnome	MS and NMR data: filtering, feature detection, alignment, multivariate data analysis
Markerlynx	Waters	LC-MS data: feature detection, alignment, PCA
MarkerView	Applied Biosystems	LC-MS data: feature detection, alignment, PCA
MassHunter	Agilent Technologies	LC-MS data: detection/extraction, alignment
Metabolic Profiler	Bruker Daltonic	MS and NMR data: data bucketing (retention time, *m/z* values), PCA
Metabolyzer	**Metabolon**	Automated chromatographic peak alignment tools. *Used as part of proprietary data processing system in commercial studies*
metAlign	PlanResearch International	LC-MS, GC-MS data: Filtering, baseline correction, feature detection, alignment
Phenomenome Profiler	Shimadzu /Phenomenome Discoveries	MS data: feature detection, alignment, statistical analysis

B)

Software	Data Type	Capabilities
AMDIS	GC-MS	Automatically extracts pure component mass spectra from highly complex GC-MS data files. Uses these purified spectra to search a mass spectral library
Chrompare	GC-FID	Feature list and raw chromatogram comparison, data normalization
COMSPARI	LC-MS, GC-MS	Visualization capabilities for investigating differences between two runs
MathDAMP	LC-MS, GC-MS, CE-MS	Direct data comparison (no feature detection), binning, baseline subtraction, smoothing, normalization
MetaboliteDetector	GC-MS	Chromatogram compression, peak detection, chromatogram deconvolution, alignment, compound integration / quantification
MET-IDEA	LC-MS, GC-MS, CE-MS	Ion intensity data extraction – outputs list of *m/z* retention time values

MSFACTs	LC-MS, GC-MS	Automated import, alignment and comparison of feature lists or raw chromatograms. Allows rapid visualization and interrogation of data
mSPECS	GC-MS, LC-MS	Management and editing of mass spectral libraries and associated information. Capabilities for the automatic management of libraries, plus automatic calculation of formulas and masses
MZMine /MZMINE2	LC-MS, GC-MS	Filtering, feature detection, alignment, normalization, visualization. MZmine 2 can process unit mass resolution and exact mass resolution data (e.g. FTMS), including fragmentation data (continuum or centroided)
XCMS/ XCMS2	LC-MS, GC-MS	Filtering, feature detection, alignment, visualization

Peak picking

Con el fin de comparar, de forma global, los metabolitos entre diferentes muestras, la detección de los picos cromatográficos (peak picking) y su correcto emparejamiento (*peak matching* o *grouping*) es crucial. Por tanto, el primer paso para el pre-procesamiento de datos procedentes de LC-MS o GC-MS se centra en distinguir los picos cromatográficos (es decir, que provienen de metabolitos) de otros picos y, sobretodo, del ruido de fondo. Este proceso, llamado *peak picking* o *peak detection*, se puede llevar a cabo mediante un proceso de deconvolución o mediante la búsqueda de picos para cada XIC extraído en todo el rango de masas aplicado. Estos XIC son creados con una determinada ventana de masa que depende de la exactitud/resolución del espectrómetro de masas en cuestión. Para ambos casos, deconvolución o XIC, existen infinidad de algoritmos disponibles que pueden ser aplicados en función del tipo de datos obtenidos (principalmente el tipo de técnica y la instrumentación utilizada). Sin embargo, aunque los algoritmos de deconvolución facilitan enormemente la interpretación de los datos generados al permitir relacionar los espectros con compuestos puros, éstos son por lo general menos sensibles (menor capacidad de detección de picos pequeños), especialmente cuando se trabaja con matrices complejas. Teniendo en cuenta lo mencionado anteriormente, el segundo método ha sido utilizado en el desarrollo de esta tesis pues, aunque representa una

mayor complejidad de datos por su colinealidad, es más poderoso en cuanto a detección se refiere.

Peak alignment

Típicamente, los estudios de *metabolite fingerprint* suelen involucrar cientos de muestras pertenecientes a distintas clases. En estos casos, es común observar derivas en los tiempos de retención de los metabolitos. Este problema es más agudo en el caso de LC y suele ser debido a (i) pequeños cambios en la temperatura, presión o en la fase móvil (composición o el flujo) (ii) cambios en la fase estacionaria de la columna (saturación/degradación) o (iii) efectos matriz de las muestras y variaciones en el extracto. Esta deriva instrumental debe ser corregida pues puede dar lugar a problemas durante el agrupamiento de picos, i.e. al emparejar cada *feature* presente en las diferentes muestras.

La corrección del tiempo de retención puede llevarse a cabo antes del proceso de peak picking (alineamiento de los perfiles cromatográficos) o a posteriori mediante un emparejamiento de picos conocidos presentes en la mayoría de muestras y a lo largo de todo el cromatograma (alineamiento de picos o peak alignment). A su vez, el alineamiento de picos puede realizarse en base a estándares internos añadidos a las muestras (compuestos exógenos) o mediante el uso de compuestos endógenos conocidos o presentes en la mayoría de muestras (Peters, van Velzen et al. 2009).

Peak integration

Los picos encontrados deben posteriormente ser integrados con el fin de permitir la comparación de sus abundancias relativas y poder así elucidar diferencias entre grupos de muestras. Obviamente, para la correcta integración, es necesario que previamente los picos hayan sido correctamente identificados y agrupados a lo largo de las muestras.

<u>Normalización</u>

La normalización de los datos permite una cuantificación más exacta de los diferentes *features* encontrados a lo largo de las muestras. El objetivo es la eliminación de cambios indeseados en la intensidad de los iones/compuestos entre los diferentes análisis. Estos cambios corresponden generalmente a deriva instrumental, debida principalmente a la sucesiva contaminación de la interfase de ionización, y dependen en gran medida del tipo de muestra, el tamaño del lote y del tipo de instrumentación utilizada. Al igual que sucede con el alineamiento de tiempos de retención, la normalización suele hacer uso de estándares internos (añadidos después de la extracción de la muestra) o de modelos estadísticos que permiten escalar los efectos de deriva para las diferentes muestras en base al set de datos completo de las muestras (Karpievitch, Dabney et al. 2012).

Como ya se mencionó en la sección de discusión de resultados del capítulo anterior, una herramienta de gran utilidad es el uso de QCs. Para este tipo de estudios, los QCs suelen ser una mezcla proporcional de todas las muestras utilizadas que es inyectada repetidamente al inicio del *batch* y periódicamente cada un determinado número de muestras (típicamente cada 5 o 10 inyecciones). Los QCs presentan un triple propósito; (i) optimizar las variables instrumentales a priori y acondicionar la columna cromatográfica, (ii) comprobar la calidad y fiabilidad de los datos obtenidos mediante análisis estadístico, (iii) permitir la evaluación de la necesidad de normalizar las muestras y, si procede, el correcto desempeño de la misma.

4.1.2 Análisis multivariante

El término "Análisis Multivariante" engloba un tipo de técnicas estadísticas para el análisis simultáneo de múltiples variables. Estas técnicas además proporcionan herramientas para interpretar la tabla de datos (por ejemplo, identificar o resaltar relaciones entre observaciones y variables, observaciones que se desvían del comportamiento esperado o tendencias deseadas e indeseadas en los datos), así

como para modelar diferencias entre grupos de observaciones y clasificar observaciones desconocidas.

Típicamente estas técnicas se clasifican entre métodos supervisados y no-supervisados. Estos últimos se caracterizan por tratar los datos sin ningún tipo de información a priori acerca de la clasificación de las muestras. Ésta es, por tanto, una buena manera de revelar agrupaciones o tendencias en los datos. Los métodos no-supervisados más comúnmente utilizados en el campo de la metabolómica son el Analisis de Componentes Principales (*Principal Component Analysis*, PCA), *hierarchical clustering* y *k-means clustering*. Por el contrario, los métodos supervisados utilizan la información relacionada con el set de datos (clases, parámetros biológicos/experimentales, etc.) para construir un modelo predictivo a partir del set de entrenamiento. Dicho modelo es posteriormente usado para predecir la clase de nuevas observaciones no incluidas en el set inicial de entrenamiento. De entre las típicas técnicas de análisis multivariante supervisadas se incluye *Partial Least Squares (PLS) regression, ortogonal-PLS (OPLS), discriminant Analysis (DA)* y *PLS Discriminant Analysis (PLS-DA)*. La elección de la técnica a utilizar depende del objetivo final y del tipo de set de datos (Want 2009).

<u>PCA</u>

Es la técnica quimiométrica más ampliamente utilizada para el análisis de datos metabolómicos basados en MS. Resulta una herramienta muy útil para obtener una visión general de sets de datos enormes (como suele ser el caso), visualizar similitudes y diferencias entre observaciones y adquirir conocimiento sobre los metabolitos responsables de los patrones y agrupaciones encontrados. Una de las ventajas de esta técnica es que puede lidiar con matrices de datos incompletas y ruidosas que contengan más variables que muestras, así como con variables altamente correlacionadas, lo cual es muy común en este tipo de datos.

PCA crea un modelo que se aproxima a los datos originales tanto como sea posible con solo unos pocos componentes principales (PCs) no-correlacionados (ortogonales). Dicho modelo se ajusta en base a aquellos PCs que explican en mayor

grado posible la varianza del set de datos, reduciendo así la dimensionalidad de éstos mientras se mantiene el máximo de información. De dicho modelo se derivan dos matrices, scores y loadings, las cuales se utilizan para producir sendos gráficos. En la figura 4.A. se puede observar un ejemplo del potencial del PCA para el tratamiento de este tipo de datos metabólicos. En dicha figura se observa la correcta agrupación de dos clases de muestras (G1 y G2) y los metabolitos significativamente influyentes en dicha agrupación.
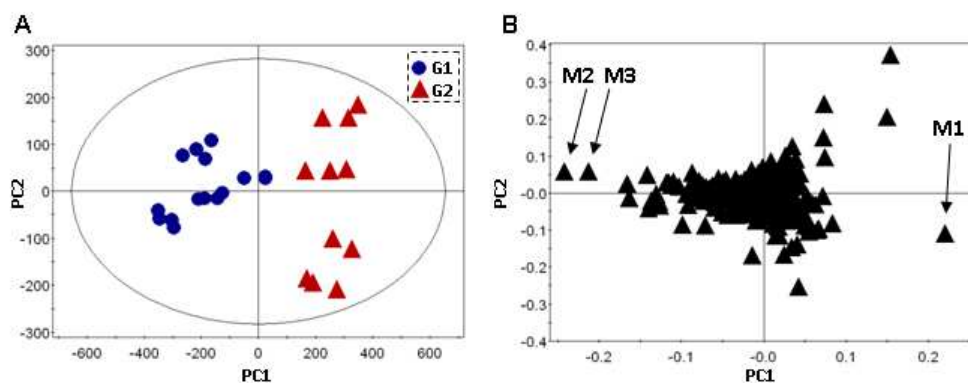

Figura 4.A. PCA correspondiente a 26 perfiles metabólicos obtenidos mediante análisis UPLC-MS de muestras de suero de rata provenientes de un estudio toxicológico. (A) PC1 vs PC2 scores plot resaltado por clases, G1 and G2; (B) PC1 vs PC2 loading plot. Metabolito M1 se encuentra sobreexpresado en G2 mientras que los metabolitos M2 y M3 presentan una mayor respuesta en G1 (Want 2011).

Métodos supervisados: PLS-DA y OPLS-DA

PCA es un excelente punto de partida para el análisis multivariante de los datos, aportando una excelente visión general de las similitudes y diferencias entre muestras. Cuando se puede observar una separación clara entre los grupos investigados, PCA certifica una correcta clasificación entre los grupos dado que la separación se realiza sin ningún tipo de información de clase incluida en el algoritmo. Sin embargo, los componentes principales modelan las variaciones máximas en la matriz de datos y estas direcciones pueden no coincidir con la separación máxima entre clases. Otras direcciones podrían ser más pertinentes para la discriminación entre los grupos de muestras objetivo (Figura 4.B). Con este fin, aproximaciones supervisadas, como PLS-DA y OPLS-DA pueden ser más adecuadas. En este caso, el algoritmo tiene

en cuenta la información de clases e intenta construir un modelo capaz de predecir dicha agrupación. En realidad, PLS-DA es un método de clasificación basado en la regresión PLS, la cual puede interpretarse como una extensión del PCA (Worley and Powers 2013).

PLS modela la asociación entre una matriz de datos X y una de respuestas Y. La matriz de datos X correspondería a la obtenida mediante el análisis de las muestras, i.e. en el caso que nos ocupa, la tabla con los datos extraídos del peak-picking de todas las muestras. Por otra parte, la matriz Y contiene información de las muestras, pudiendo ser esta información una variable continua (como por ejemplo calificación de la muestra, tiempo trascurrido desde la infección, pH, etc.) o categórica (denominación de origen, tipo de tratamiento, etc.). Por tanto, PLS-DA es a su vez una extensión de PLS para variables categóricas o puntuales. Más específicamente, mientras que PCA es una proyección de la varianza máxima en la matriz de datos X, PLS (y por tanto, PLS-DA) genera un modelo de covarianza máxima entre X y Y.

PLS-DA resulta una técnica muy apropiada para análisis metabolómico de datos de MS dadas las características de la matriz X las cuales suelen presentar un considerablemente mayor número de variables que de observaciones. Además, PLS-DA no se ve afectada con variables fuertemente correlacionadas, como es común en este tipo de datos (isótopos, fragmentos o compuestos relacionados). Sin embargo, los modelos PLS-DA se ven negativamente afectados por variaciones sistemáticas en la matriz X no relacionadas con la matriz Y, como por ejemplo una elevada varianza en las diferentes clases. Cuando la dirección preferencial en la separación de clases en PLS-DA es una combinación de PCs, los modelos pueden resultar difícilmente interpretables ya que la variación entre clases y dentro de las clases es fusionada. Una modificación del método PLS, llamado OPLS (Orthogonal PLS), se presenta como una excelente alternativa para facilitar la interpretación.

OPLS (así como OPLS-DA para variables categóricas) separa la variación en la matriz de datos X en dos partes: una correlacionada con Y y otra ortogonal a Y. Esta partición facilita la interpretación del modelo ya que ambas variaciones (entre clases y dentro de las clases) son separadas (Figura 4.B.).
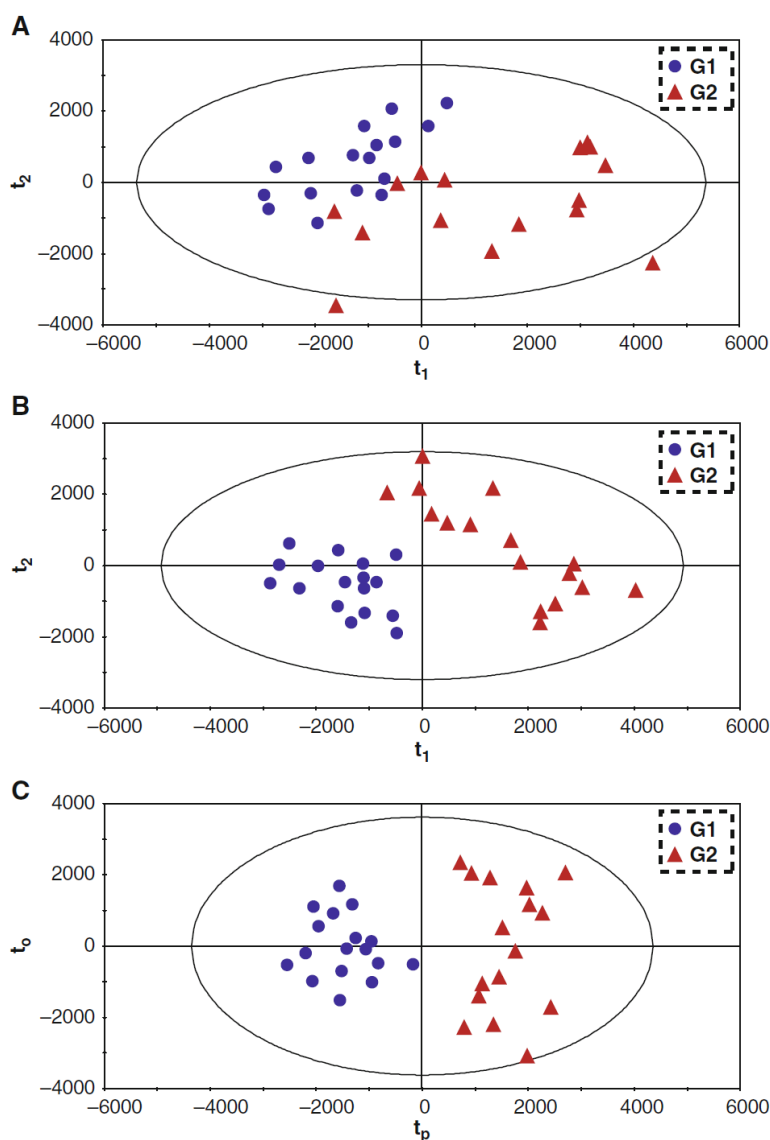
Figura 4.B. Scores Plot (PC1 vs. PC2) de datos obtenidos del análisis de suero de rata modelados mediante (a) PCA, (b) PLS-DA y (c) OPLS-DA. La separación entre dos grupos de animales se puede observar directamente con PCA. Esta separación se amplifica al aplicar PLS-DA puesto que el algoritmo tiene en cuenta ambas clases para construir el modelo. Mediante OPLS-DA, el PC1 es usado como componente predictivo de las clases, facilitando la interpretación del modelo y la identificación de marcadores (Want 2011).

---------------------- O -----------------------

En este **capítulo,** se explora el análisis metabolómico para la identificación de marcadores fiables que puedan ser utilizados para asegurar la autenticidad de muestras de alimentos. En el **artículo científico 5** se realiza un análisis untargeted de muestras de naranjas de diferentes variedades, cosecha y procedencia con la finalidad de obtener un marcador fiable para la distinción de la procedencia de naranjas de variedades de maduración tardía, con especial énfasis en la diferenciación de muestras procedentes de la Comunidad Valenciana en comparación con frutos cultivados en el hemisferio sur.

En el segundo trabajo de este capítulo, **artículo científico 6**, se busca ahondar en el análisis metabolómico untargeted para la caracterización de vinos de diferente denominación de origen. Para ello, se comparan muestras de vino de 3 importantes áreas vinícolas de España mediante dos plataformas diferentes donde se combinan Orbitrap y (Q)TOF MS, así como aproximaciones distintas. Este artículo muestra la viabilidad de ambos aproximaciones, mostrando la capacidad de ambas para encontrar marcadores capaces de realizar la clasificación inequívoca del origen de la muestra. Además se exaltan las similitudes y diferencias de ambas aproximaciones.

4.2. Artículo científico 5

"Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry"

R. Díaz, O. J. Pozo, J. V. Sancho and F. Hernández.

Food Chemistry, 157 (2014) 84-93.

Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry

Ramon Díaz [a], Oscar J. Pozo [b], Juan V. Sancho [a,*], Félix Hernández [a]

[a] Research Institute for Pesticides and Water, University Jaume I, Av. Sos Baynat S/N, 12071 Castellón, Spain
[b] Bioanalysis and Analytical Services Research Group, Neuropsychopharmacology Research Program, IMIM-Hospital del Mar, Barcelona, Spain

A B S T R A C T

In this work, hybrid quadrupole time-of-flight mass spectrometer (QTOF MS) coupled to ultra high performance liquid chromatography (UHPLC) has been used for biomarkers identification for correct authentication of Valencia (Spain) oranges. Differentiation from foreign Argentinean, Brazilian and South African oranges has been carried out using XCMS application and multivariate analysis to UHPLC-(Q)TOF MS data acquired in both, positive and negative ionisation modes. Several markers have been found and corroborated by analysing two seasons samples. A seasonal independent marker was found and its structure elucidated using accurate mass data and MS^E fragmentation spectrum information. Empirical formula was searched in Reaxys database applying sub-structure filtering from the fragments obtained. Three possible structures were found and citrusin D, a compound present in sweet oranges, has been identified as the most plausible as it fits better with the product ion scan performed for this compound. As a result of data obtained in this work, citrusin D is suggested as a potential marker to distinguish the geographic origin of oranges.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Food quality and traceability are issues of wide concern to costumers and markets (Arvanitoyannis & Vaitsi, 2007; Reid, O'Donnell, & Downey, 2006). This quality is often linked to their origin as, for example, French wine or Spanish ham, which in some cases increase considerably the price of the product (Luykx & van Ruth, 2008). Thus, to guarantee the traceability of these items becomes essential, not only for costumers but also for manufacturers and dealers (Reid, O'Donnell, & Downey, 2006). To this aim, analytical methodologies able to unequivocally distinguish and detect the fraud are welcome.

Although not an especially high value product, oranges are an important market in the Valencian region, located in the east coast of Spain. This "Denomination of Origin" (D.O.) is well-known for Spanish costumers and the region economy's strongly depends on it, as recently published (Malet, 2010). Furthermore, it is suspected that a fraud really exists, mainly in late maturing orange varieties in order to overcome the production limitations during the summer, when no oranges are available at Valencia latitude.

Thus, it is not uncommon that oranges are imported from southern hemisphere countries during this period and sold with Valencian D.O. Unfortunately, there is a lack on reliable analytical methodologies to avoid this fraud.

Normally, food authenticity methods are focused on the determination of a few target compounds that characterise the sample, as some metals by Inductively Coupled Plasma (ICP) (Anderson & Smith, 2006; Perez, Smith, & Anderson, 2006), natural products as amino acids (Gómez-Ariza, Villegas-Portero, & Bernal-Daza, 2005) and antioxidants (Iijima, Suda, Suzuki, Aoki, & Shibata, 2008; Makris, Kallithraka, & Mamalos, 2006) or even on searching for common adulteration substances, as tartaric acids in orange juice (Saavedra, García, & Barbas, 2000). One of the most promising approaches to solve these problems is metabolomics, or more precisely the metabolic fingerprinting (Lu et al., 2008), defined as "*the unbiased, global screening approach to classify samples based on metabolite patterns or "fingerprints" that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites*" (Dettmer, Aronov, & Hammock, 2007). Although typically focussed on human (or animal) tissues (liver, lung, kidney, etc.) or biological fluids (mainly urine and plasma) to evaluate the effect of drugs (Nicholson, Connelly, Lindon, & Holmes, 2002) as well as to understand and diagnose

* Corresponding author. Tel.: +34 964387363; fax: +34 964387368.
  E-mail address: sanchoj@uji.es (J.V. Sancho).

Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry

Ramon Díaz[a], Oscar J. Pozo[b], Juan V. Sancho[a],* , Félix Hernández[a]

[a]*Research Institute for Pesticides and Water, University Jaume I, Av. Sos Baynat S/N, 12071 Castellón, Spain*

[b]*Bioanalysis and Analytical Services Research Group, Neuropsychopharmacology Research Program, IMIM-Hospital del Mar, Barcelona, Spain*

HIGHLIGHTS

• We performed an untargeted metabolic fingerprint for oranges origin differentiation.
• UHPLC-(Q)TOF MS data have been processed using XCMS and multivariate analysis.
• Results have been validated by increasing the sample number and additional season and origin.
• Citrusin D has been tentatively identified has an optimal marker for the geographical authentification

ABSTRACT

In this work, hybrid quadrupole time-of-flight mass spectrometer (QTOF MS) coupled to ultra-high performance liquid chromatography (UHPLC) has been used for biomarkers identification for correct authentication of Valencia (Spain) oranges. Differentiation from foreign Argentinean, Brazilian and South African oranges has been carried out using XCMS application and multivariate analysis to UHPLC-(Q)TOF MS data acquired in both, positive and negative ionisation modes. Several markers have been found and corroborated by analysing two seasons samples. A seasonal independent marker was found and its structure elucidated using accurate mass data and MSE fragmentation spectrum information. Empirical formula was searched in Reaxys database applying sub-structure filtering from the fragments obtained. Three possible structures were found and citrusin D, a compound present in sweet oranges, has been identified

as the most plausible as it fits better with the product ion scan performed for this compound. As a result of data obtained in this work, citrusin D is suggested as a potential marker to distinguish the geographic origin of oranges.

1. INTRODUCTION

Food quality and traceability are issues of wide concern to costumers and markets (Arvanitoyannis and Vaitsi, 2007 and Reid et al., 2006). This quality is often linked to their origin as, for example, French wine or Spanish ham, which in some cases increase considerably the price of the product (Luykx & van Ruth, 2008). Thus, to guarantee the traceability of these items becomes essential, not only for costumers but also for manufacturers and dealers (Reid, O'Donnell, & Downey, 2006). To this aim, analytical methodologies able to unequivocally distinguish and detect the fraud are welcome.

Although not an especially high value product, oranges are an important market in the Valencian region, located in the east coast of Spain. This "Denomination of Origin" (D.O.) is well-known for Spanish costumers and the region economy's strongly depends on it, as recently published (Malet, 2010). Furthermore, it is suspected that a fraud really exists, mainly in late maturing orange varieties in order to overcome the production limitations during the summer, when no oranges are available at Valencia latitude. Thus, it is not uncommon that oranges are imported from southern hemisphere countries during this period and sold with Valencian D.O. Unfortunately, there is a lack on reliable analytical methodologies to avoid this fraud.

Normally, food authenticity methods are focused on the determination of a few target compounds that characterise the sample, as some metals by Inductively Coupled Plasma (ICP) (Anderson and Smith, 2006 and Perez et al., 2006), natural products as amino acids (Gómez-Ariza, Villegas-Portero, & Bernal-Daza, 2005) and antioxidants (Iijima et al., 2008 and Makris et al., 2006) or even on searching for common adulteration substances, as tartaric acids in orange juice (Saavedra, García, & Barbas, 2000). One of the most promising approaches to solve these problems is metabolomics, or more precisely the metabolic fingerprinting (Lu et al., 2008), defined as "the unbiased, global screening approach to classify samples based on metabolite patterns or "fingerprints" that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites" (Dettmer, Aronov, & Hammock, 2007). Although typically focussed on human (or animal) tissues (liver, lung, kidney, etc.) or biological fluids (mainly urine and plasma) to evaluate the effect of drugs (Nicholson, Connelly, Lindon, & Holmes, 2002) as well as to understand and diagnose diseases (Pendyala, Want, Webb, Siuzdak, & Fox, 2007), metabolomic approaches have been recently applied to food characterisation for

coffee (Choi et al., 2010 and Risticevic et al., 2008), wine (Arvanitoyannis et al., 1999 and Cuadros-Inostroza et al., 2010), tomato (Arvanitoyannis and Vaitsi, 2007 and Consonni et al., 2009) or fruit juices (Vaclavik, Schreiber, Lacina, Cajka, & Hajslova, 2011) among others.

There are a wide range of chromatographic modes and detector combinations (Arvanitoyannis & Vaitsi, 2007) available to achieve this global screening, like UV–VIS or NIR spectroscopy, etc. (Cozzolino, Smyth, & Gishen, 2003), although two main tools are used NMR spectroscopy and gas/liquid chromatography coupled to mass spectrometry (GC–MS and LC–MS) (Cevallos-Cevallos, Reyes-De-Corcuera, Etxeberria, Danyluk, & Rodrick, 2009). NMR has been the technique of choice during the last years due to its universality and versatility as well as the stability and large structure information obtained. In the last years, MS hyphenated to chromatography is appearing as a less expensive and more sensitive alternative approach (Dunn & Ellis, 2005), very powerful when modern analyzers are employed.

Hybrid quadrupole time-of-flight (Q-TOF) or Orbitrap are the MS analysers recently used because of their high resolution (HR), mass accuracy and full-spectrum acquisition capabilities. As well-known, GC–MS is applied to (semi)volatile and no thermo-labile compounds and it presents a more restricted application field (Dunn & Ellis, 2005) due to the unavailability of GC systems coupled to state-of-the-art HRMS. However, the analysis of volatiles compounds by head space solid phase micro-extraction (HS-SPME) has been also successfully applied for the origin classification (Montero-Prado, Bentayeb, & Nerín, 2013). On the other hand, the less complex sample preparation typically required by LC–MS methods, together with the wider range of separation mechanism (i.e. reversed phase, HILIC, etc.) in liquid chromatography, makes this technique more suitable to develop metabolomic profiling methods and expand their applicability from medium to highly polar compounds. In addition, recent advances in chromatographic separations thanks to reducing the particle size of the stationary phase sorbent (i.e. ultra high-performance liquid, UHPLC), facilitates data processing (Moco, Vervoort, Bino, & De Vos, 2007), especially in the peak picking process where peak shape and resolution are crucial.

In this work, orange samples from Valencia Spanish region and foreign samples (Argentina, South Africa and Brazil) have been investigated using multivariate analysis from the data obtained by UHPLC-QTOF MS with the objective to discover useful biomarkers for origin differentiation. QTOF MS has been used under MS^E mode, i.e. simultaneous acquisition at low (LE) and high collision energy (HE), which provides useful information on the (de)protonated molecules (commonly at LE) and on the main fragments ions (commonly in HE) (Díaz, Ibáñez, Sancho, & Hernández, 2011). This approach provides helpful information for the speed-up identification of the discovered markers.

2. MATERIALS AND METHODS

2.1 Reagents and chemicals

Around 40 analytes were studied for performance optimisation and a total of 230 compounds were included in the empirical mass spectral library (see Table 1). Reference compounds were purchased from Acros Organics (Geel, Belgium), Bayer Hispania (Barcelona, Spain), Fort Dodge Veterinaria (Gerona, Spain), Vetoquinol Industrial (Madrid, Spain), Aventis Pharma (Madrid, Spain), Sigma Aldrich (St Louis, MO, USA), Cerilliant (Round Rock, TX, USA), Dr. Ehrenstorfer (Augsburg, Germany), Riedel-de Haën (Seelze, Germany), the National Measurement Institute (Pymble, Australia) and Fluka (Buchs, Switzerland). All reference materials had purities higher than 93% (w/w).

2.2. Instrumentation

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (Q-oaTOF Premier, Waters Micromass, Manchester, UK), using an orthogonal Z-spray-ESI interface operating in positive and negative ion mode. The UHPLC separation was performed using an Acquity UPLC BEH C18 1.7 μm particle size analytical column 100 × 2.1 mm (Waters) at a 300 μL/min flow rate. The mobile phases used were A = H2O with 0.01% HCOOH and B = MeOH with 0.01% HCOOH. The percentage of organic modifier (B) was changed as follows: 0 min, 10%; 1.5 min, 10%; 16 min, 90%; 18 min, 90%; 18.01 min, 10% in a total run time of 20 min for the positive ionisation mode and 0 min, 10%; 2.5 min, 10%; 16.5 min, 75%; 18 min, 90%; 20 min, 90%; 20.01 min, 10% in a total run time of 22 min for the negative ionisation mode. Nitrogen was used as the drying gas and nebulizing gas. The desolvatation gas flow was set at 600 L/h. TOF-MS resolution was approximately 10.000 at full width half maximum (FWHM) in V-mode at m/z 556.2771. MS data were acquired over an m/z range of 50–1000. A capillary voltage of 3.5 kV and 3.0 kV for positive and negative ion modes, respectively, and cone voltage of 25 V were used. Collision gas was argon 99.995% (Praxair, Valencia, Spain). The interface temperature was set to 350 °C and the source temperature to 120 °C. The column temperature was set to 60 °C.

For MSE experiments, two acquisition functions with different collision energies were created. The first one, the low energy function (LE), selecting a collision energy of 4 eV, and the second one, the high energy (HE) function, with a collision energy ramp ranging from 15 to 40 eV in order to obtain a greater range of fragment ions. The LE and HE functions settings were for both

a scan time of 0.2 s and an inter-scan delay of 0.05 s. The automated attenuated function was also selected to correct for possible peak saturations (extended mode).

Calibrations were conducted from m/z 50 to 1000 with a 1:1 mixture of 0.05 M NaOH:5% HCOOH diluted (1:25) with acetonitrile:water (80:20) and spiked with fungicide imazalil at a final concentration of 500 μg/L, at a flow rate of 10 μL/min. For automated accurate mass measurement, the lock-spray probe was used, using as lock mass a solution of Leucine-enkephalin (2 μg/mL) in acetonitrile:water (50:50) at 0.1% HCOOH pumped at 30 μL/min through the lock-spray needle. A cone voltage of 65 V was selected to obtain adequate signal intensity for this compound (□500 counts). The (de)protonated molecule of Leucine-enkephalin, at m/z 556.2771 in positive mode and m/z 554.2615 in negative mode, was used for recalibrating the mass axis and ensuring a robust accurate mass measurement along time. It should be noted that all the exact masses shown in this work have a deviation of 0.55 mDa from the "true" value, as the calculation performed by the MassLynx software uses the mass of hydrogen instead of a proton when calculating [M+H]+ exact mass. However, because this deviation is also applied during mass axis calibration, there is not negative impact on the mass errors presented in this article. MS data were acquired in centroid mode.

2.3. Samples

Orange samples of different varieties and origin were obtained from a manufacturer company in the Valencia region. Concretely, 15 samples of Valencia Origin and 9 foreign samples were initially analysed. In a second phase, the number of samples was widened using next year oranges, collecting a total of 42 samples: 24 from Valencia and 18 from abroad. All the foreign samples were Valencia variety (VA) while the samples from Valencia ones Valencia (VA), Navelina (NV), Navel-late (NL) and Valencia-late (VL) varieties, all of them late maturing crops. The samples consisted on at least 6 oranges from the same variety, origin and location. All the samples were treated equally after the harvest process.

2.4. Sample treatment

Orange samples were equally handled during all the analytical process. They were stored at −18 °C until analysis and let thaw during 12 h at room temperature before processing. The whole samples, including peel and pulp, were chopped and homogenised in Homogeniser Thermomix TM30 (Vorwerk, Madrid, Spain) at room temperature during 2 min. Then, 10 g of sample was extracted with a 20 mL water:MeOH (50:50) in a mechanical shaker during 90 min.

Afterwards, it was centrifuged at 5000 rpm for 10 min and the supernatant was taken. In order to reduce the organic percentage, a fivefold dilution with water was performed. The diluted sample extract (25 μL) was injected into the UPLC-QTOF MS system in both positive and negative ionisation modes, randomly in order to avoid instrumental drift effect on the results. Quality Control (QC) samples were included every 5 injections for analysis control purposes. QC's consisted on a representative average samples formed by a pool of 1 mL-aliquots of different final sample extracts.

2.5. Data processing

UPLC-(Q)TOF MS data were converted from proprietary to netCDF format using Databridge application (within MassLynx v 4.1; Waters Corporation) and pre-processed using XCMS free software. Two peak picking methods were tested: matched filter (S/N threshold = 3, FWHM = 15, step = 0.1) and centwave (peak width = 5–20 s, S/N threshold = 10, ppm = 25). Peak grouping was carried out using a density grouping method with an initial band width of 30 s. Then, peaks retention time was aligned using a non-linear loess based correction fitting allowing a maximum of five missing samples to consider a peak group as "well behaved" and eligible to be included as an element of the retention time correction spine.

Multivariate analysis was carried out using SIMCA-P (Umetrics, Sweden). Partial Least Square Discriminant Analysis (PLS-DA) was employed to distinguish samples according to their origin. Then, orthogonal PLS-DA (OPLS-DA) analysis was used to highlight the markers to differentiate between Valencia and foreign classes.

3. RESULTS AND DISCUSSION

3.1. First season samples

In a first step, 9 foreign (Argentina and South Africa) and 15 Valencia (Spain) orange samples were compared. Sampling process becomes problematic, mainly due to their seasonality as well as to the difficulties to have absolute sureness of their geographical origin.

3.1.1. Sample analysis

For analysis, both orange peel and pulp were chopped together and a simple extraction method with water:MeOH mixture was carried out followed by a water dilution to reduce the organic percentage and the instrumental drift (obtaining better robustness). As the goal of this

methodology was to achieve a potential marker for the discrimination of the orange samples according of their origin, not taken into account flavour or quality of the sample, no separation of the pulp and peel was performed as both could be relevant for this purpose. Additionally, in order to reduce the inherent instrumental drift, samples were injected randomly in the UHPLC-(Q)TOF MS system and sensitivity was checked by including a QC samples in every sample batch. QC sample was also used for the chromatographic system stabilization by injecting 10 initial QC replicates. TOF MS general conditions were set according to our previous knowledge of the instrument and injections were performed in both positive and negative ionisation modes. Gradient elution was optimised based on better resolving the visible peaks in the Base Peak Ion (BPI) chromatogram from the QC sample.

3.1.2. Data process

Data processing is probably one of the bottlenecks of the LC–MS based metabolomics. Chromatographic peaks have to be extracted from the total ion chromatogram (TIC), which requires specific software in order to get as much reliable peaks as possible. Furthermore, deviations in retention time during analysis as well as instrumental drift need to be corrected as the LC stability is lower compared to, for example, NMR instruments. To solve these limitations, samples were injected in a single batch, inserting Quality Control (QC) samples every five samples.

As can be seen in Fig. 1a, no appreciable differences were found in the BPI chromatograms for oranges of different origins. Then, to extract all ion information, XCMS software was used. Two peak picking methods were tested: centwave and matched filter, obtaining always better results with the later based on the number of peaks (i.e. components). 10,503 and 8619 peaks were found with Matched Filter against 4029 and 1264 obtained with centwave method, in positive and negative ionisation mode, respectively. Retention time correction and peak integration was also performed within XCMS (with the parameters explained in the experimental section).
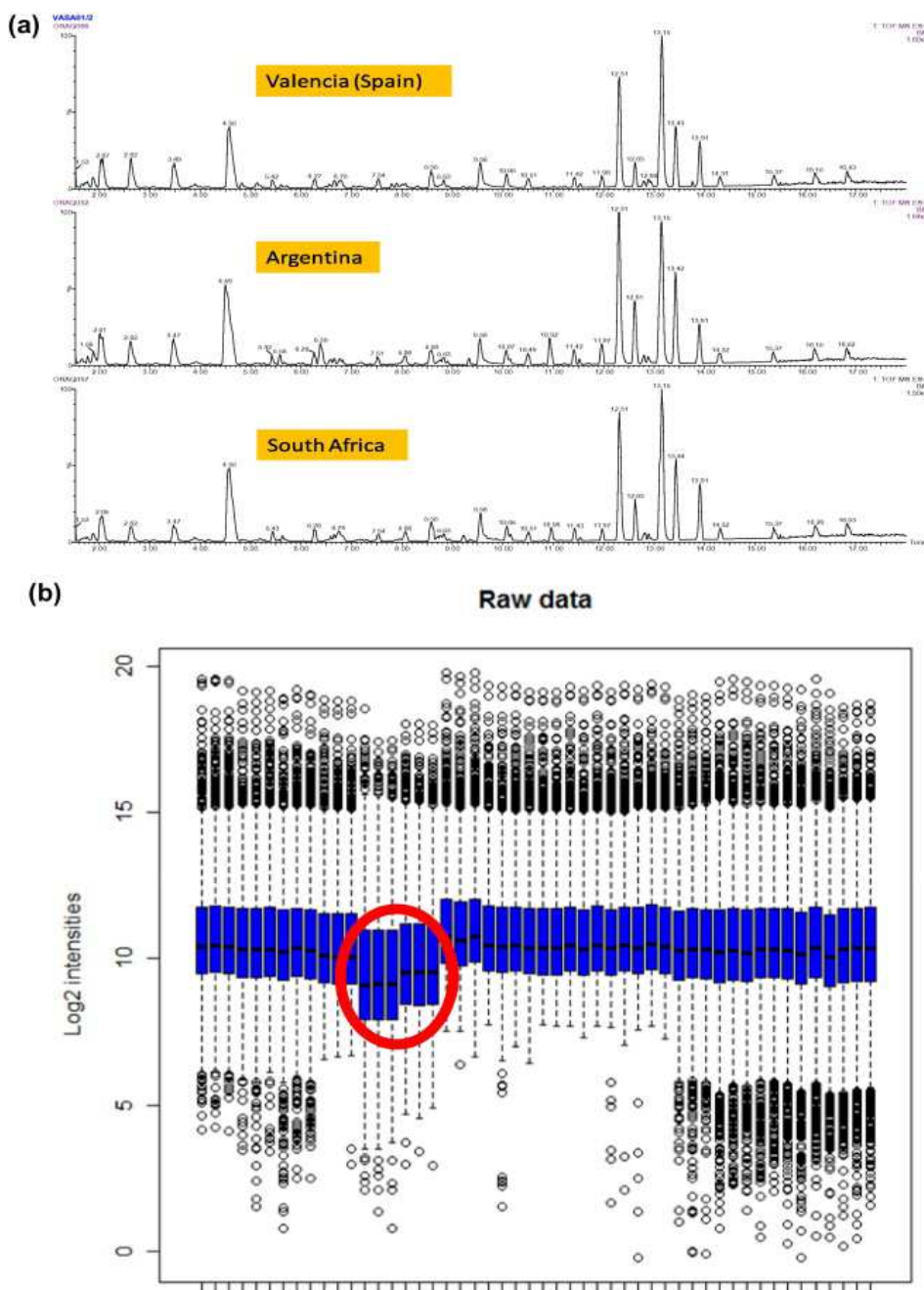
Fig. 1 (a) BPI chromatogram of Valencia, South Africa and Argentina example samples and (b) boxplot of the Log2 of the intensities of the components found in the different samples analysed in positive ESI mode. QC diluted samples (lower median) are highlighted.

Components intensity normalisation was considered to avoid possible differentiations due to water content of the sample or instrumental signal drift. However, median intensities of all the detected features in the samples (Fig. 1b) did not show significant differences (except for diluted QCs, as expected) what suggested that normalisation was not required. In fact, Principal Component Analysis (PCA) applied to raw and normalised data led to similar results (data not shown).

PLS-DA, a supervised method, was applied to construct and validate a statistical model to find differences in non-volatile compounds composition between Valencian and foreign samples. Five groups were considered: Valencia (VA), foreign (EXT) and QCs (undiluted, twofold and fourfold diluted). As can be seen in Fig. 2a, for the positive ionisation data, separation was unequivocal, not only for the main classes but also for Argentina and South Africa samples (which were considered as a single class). Furthermore, the recognition ability after cross-validation model, i.e. the model capability to correctly classify an unknown sample, was 100% for all the classes.
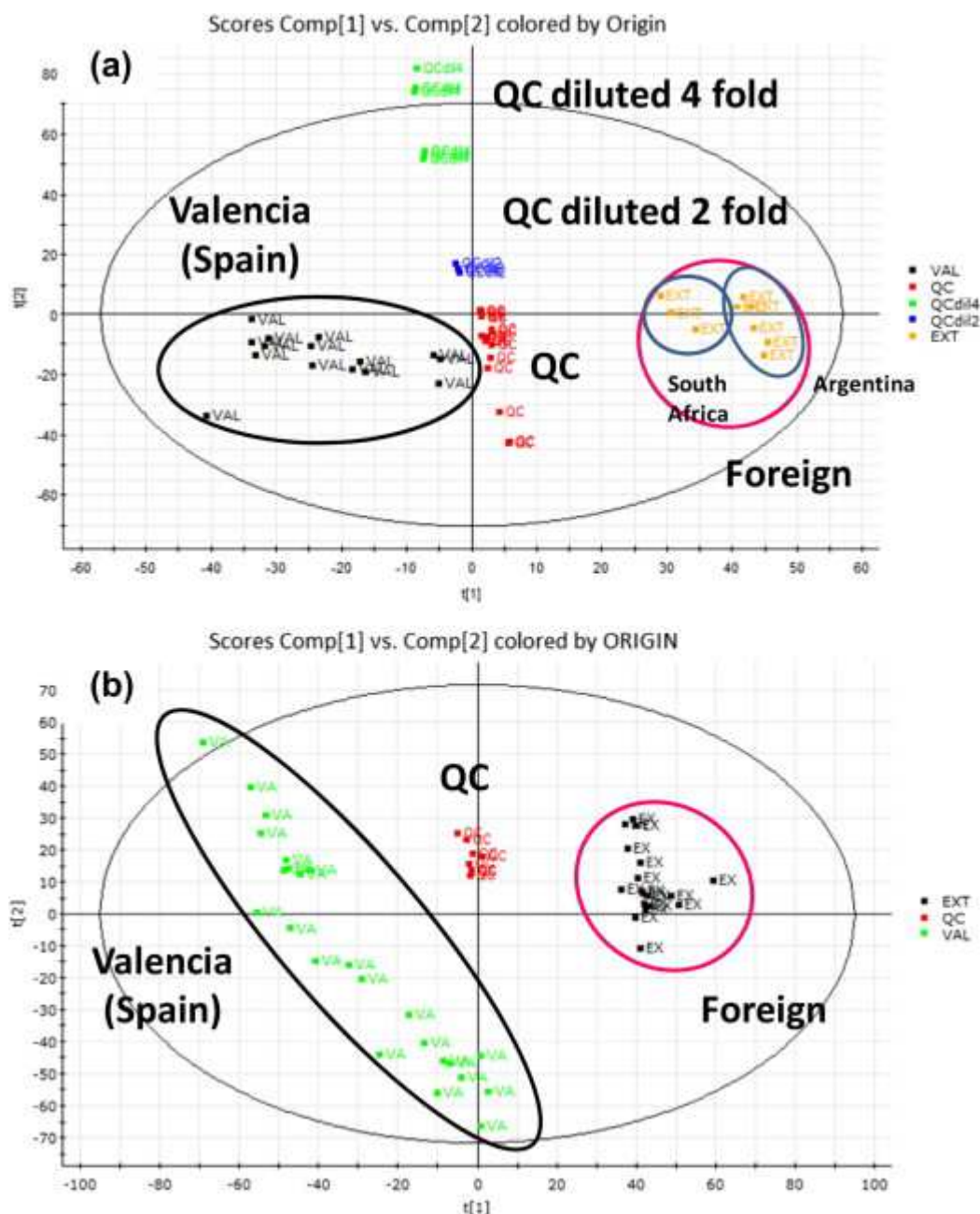
Fig. 2. PLS-DA score plot of the first two components of the non-normalised pre-processed (a) first and (b) two seasons data.

In order to easily highlight the significant biomarkers, OPLS-DA was applied to EXT and VA classes. Most relevant features (i.e. m/z–RT pairs) were selected from the S-plot (Fig. 3). The first 20 m/z–RT pairs ordered by p-value significance are shown in Table 1a. As it can be noticed for

the positive ionisation mode, the main 20 features that contribute to the separation of sample classes belong to only 5 different compounds. In fact, 12 out of 20 features were adducts and/or fragments of the same compound with RT 4.83 min (290 s). This was confirmed by our data, as the chromatographic RT and peak shape was exactly the same after performing narrow window eXtracted Ion Chromatograms (nw-XICs) of these 12 ions (example in Fig. 4a). Signal differences for this compound as a function of oranges origin were verified by depicting the nw-XICs for most abundant ions. An example is shown in Fig. 4b using m/z 343.
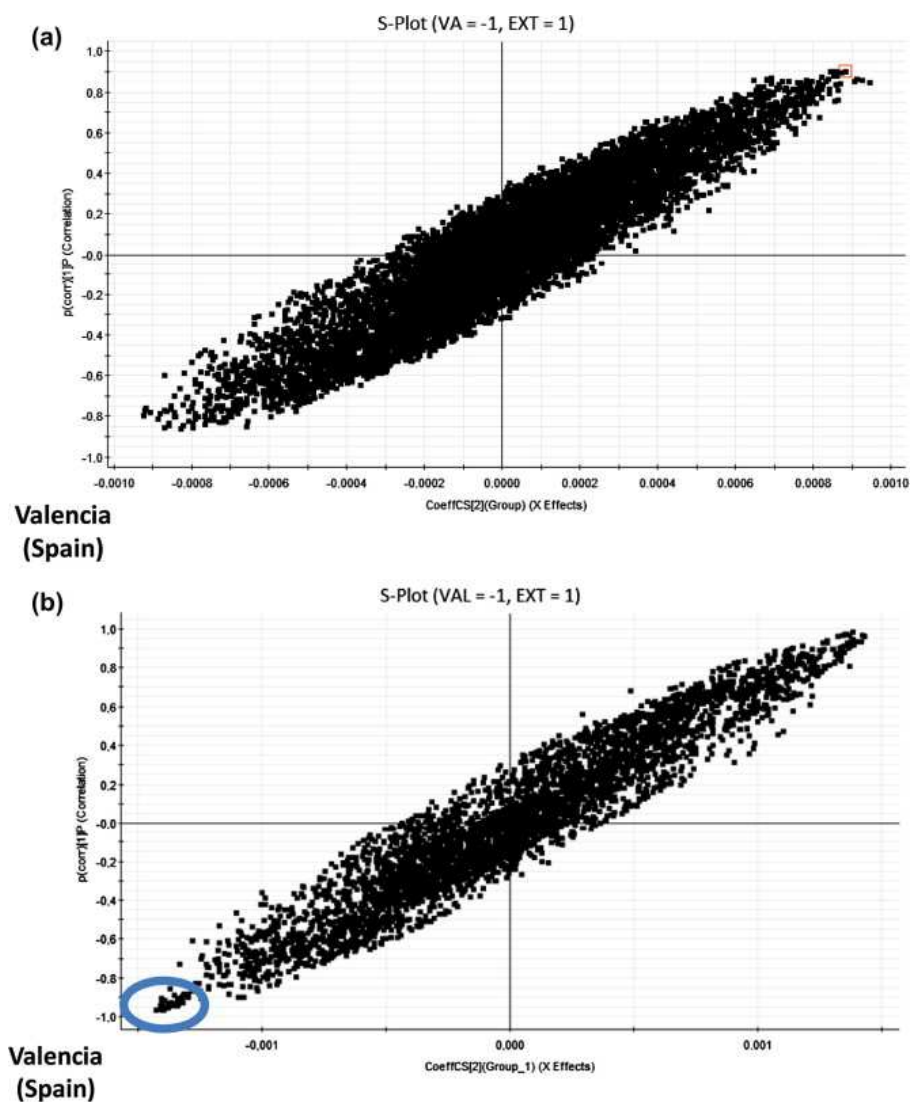
Fig. 3. OPLS-DA S-plot of classes Valencia and foreign samples of the non-normalised pre-processed data for (a) first season and (b) two seasons. More important features for Valencian samples (which correspond to RT 4.83 min compound) are highlighted.

Table 1. First 20 MFs ordered by lowest p-value (best markers candidates) for the positive ESI mode in (a) first season and (b) two season data. Fold is defined as the fold-change of the average area of VAL and EXT samples.

VIP	*m/z_*RT pair	Fold	*p*-Value	Average *m*/z (Da)	Average RT (min)
(a)					
1	M603T503[a]	2.7	6.7E−16	603.2043	8.38
2	M499T379[b]	2.1	1.8E−15	499.1216	6.31
3	M604T504[a]	3.1	4.8E−14	604.2096	8.39
4	M213T350	2.3	6.3E−14	213.1143	5.83
5	M703T290[c]	4.2	2.1E−13	703.3033	4.83
6	M709T290[c]	4.5	2.4E−13	709.2595	4.83
7	M343T290[c]	5.4	5.2E−13	343.1562	4.83
8	M131T290[c]	2.6	7.0E−13	131.0496	4.83
9	M325T290[c]	5.2	7.2E−13	325.1438	4.83
10	M707T290[c]	3.4	8.1E−13	707.2523	4.83
11	M533T290[c]	3.9	9.5E−13	533.1717	4.83
12	M534T290[c]	3.8	1.2E−12	533.6737	4.83
13	M708T290[c]	3.2	1.6E−12	708.2571	4.83
14	M163T290[c]	2.5	1.7E−12	163.0761	4.83
15	M619T503[a]	2.4	2.4E−12	619.1797	8.38
16	M483T379[b]	2.0	3.6E−12	483.1478	6.31
17	M164T290[c]	2.6	5.0E−12	164.0808	4.83
18	M313T719	3.8	5.7E−12	313.1078	11.99
19	M598T503[a]	3.2	7.8E−12	598.2515	8.38
20	M702T290[c]	4.8	1.2E−11	702.2979	4.83
(b)					
1	M132T316[a]	2.0	6.0E−15	132.0522	5.26
2	M103T315[a]	1.8	1.3E−14	103.0529	5.25
3	M131T316[a]	1.9	1.8E−14	131.0491	5.26
4	M832T531	1.8	1.2E−13	832.2372	8.85
5	M846T513	3.4	2.1E−13	846.3078	8.55
6	M801T529[b]	2.1	3.6E−13	801.2222	8.81
7	M820T509	2.5	4.1E−13	820.4190	8.48
8	M656T603[d]	1.4	4.3E−13	656.1951	10.05
9	M655T602[d]	1.4	4.7E−13	655.1887	10.04
10	M163T316[a]	1.9	3.4E−12	164.0809	5.26
11	M164T315[a]	1.9	2.8E−12	663.1986	5.25
12	M861T672	1.6	4.8E−13	861.2653	11.20
13	M511T521	2.2	5.7E−13	511.1685	8.68
14	M524T570	1.6	6.3E−13	524.1756	9.50

VIP	*m*/z_RT pair	Fold	*p*-Value	Average *m*/z (Da)	Average RT (min)
15	M802T529[b]	2.0	7.7E−13	802.2053	8.81
16	M233T673[c]	1.7	9.8E−13	233.1226	11.20
17	M234T673[c]	1.7	1.2E−12	234.1267	11.20
18	M527T292	2.2	2.0E−12	527.1725	8.78
19	M630T705	2.3	2.6E−12	630.2735	11.75
20	M816T582	1.9	3.0E−12	816.2280	9.70

[a,b,c] Features with the same RT, so belonging to the same compound (adducts, dimmers and fragments).

[a,b,c,d] Features with the same RT, so belonging to the same compound (adducts and fragments).
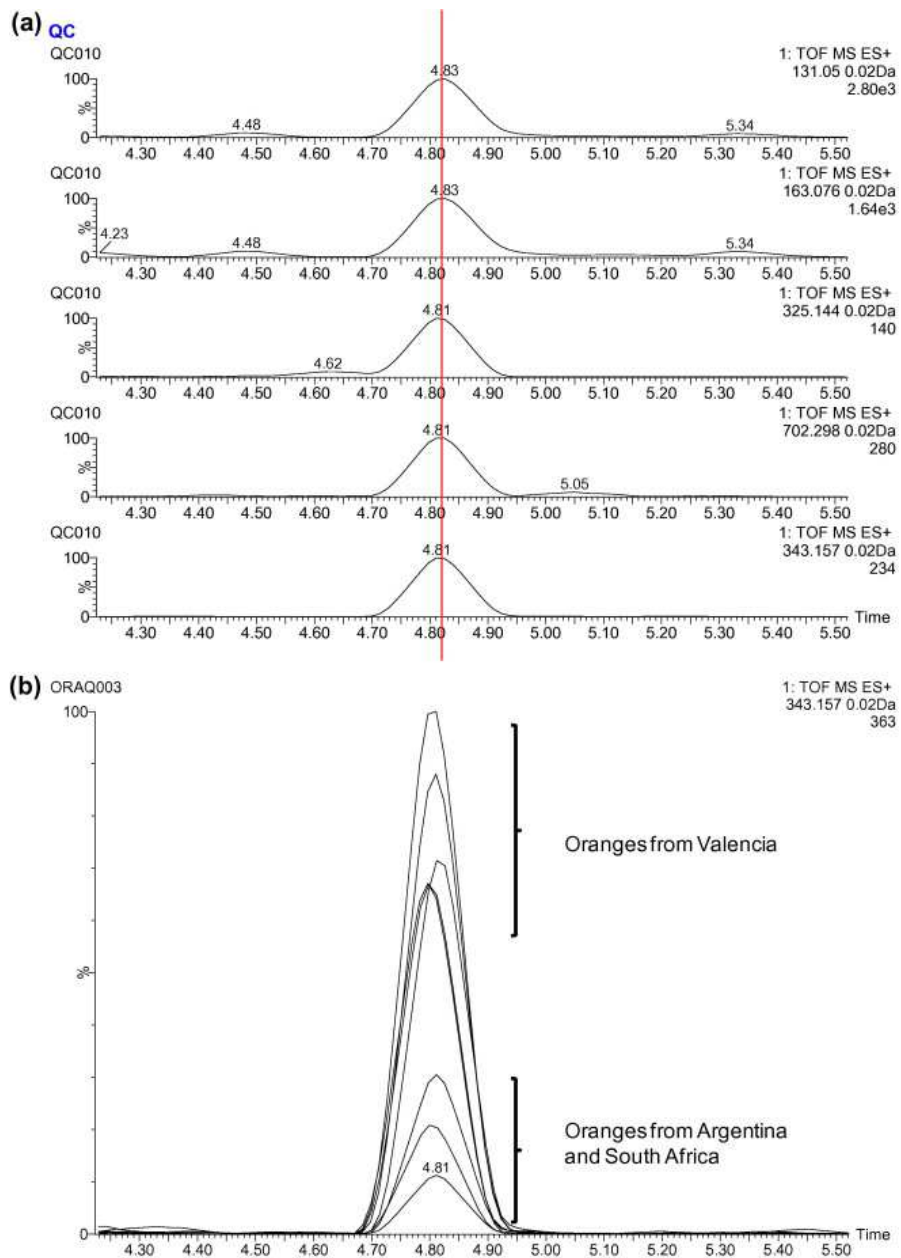
Fig. 4. 0.02 Da nw-XIC of (a) 5 of the coeluting ions in a QC sample obtained from the 20 first markers candidates list and (b) selected marker (m/z 343, RT 4.83 min) in different samples.

3.2. Widening with a second season samples

In a second phase, the same procedure was applied for a higher number of samples; samples for two seasons were joined obtaining a total number of 24 VAL and 15 EXT. In the later case, an additional south hemisphere origin, Brazil, was also included. The purpose was to confirm the validity of the biomarkers found, even for different year samples and additional origins. As can be seen in Fig. 2b, the PLS-DA model led to a clear separation of VAL and EXT samples and QCs show high reproducibility. In this case, diluted QC were not injected as it was already demonstrated, during the first season samples, that no saturation in the UHPLC-(Q)TOF MS system occurred.

The compound M603T503, seemed to be a good marker when considering only the first season data, but it was not appropriate when using the two seasons data (so it does not appear in Table 1b). Oppositely, the compound at RT 4.83 min was again found to be an optimal marker for the origin differentiation, now including Argentina, South Africa and Brazil samples. Fig. 4b shows the S-plot for the two seasons data and clearly illustrates that the marker led to the best statistical difference in the separation of the two groups. The significance of this compound is also exhibited in Table 1b with a low p-value. Thus, greater confidence on the marker ability to differentiate between orange origins was achieved.

3.3. Structural elucidation of the candidate marker

After corroborating that the compound at RT 4.83 min was found the best marker with no seasonal effect (i.e. in the two seasons samples study), elucidation was carried out using accurate mass information obtained from both TOF MS functions. The first step was to find the [M+H]+ pseudo-molecular ion as more than 10 abundant m/z ions could be already observed in the LE spectrum at this RT. Looking at the LE spectrum (Fig. 5a) two clusters could be easily observed: m/z 360, 365, 381 and 702, 707, 723. Both clusters match, within 2 mDa tolerance, with the ammonium, sodium and potassium adducts of a compound with a molecular formula $C_{16}H_{22}O_8$ ([M+NH4]+, [M+Na]+, [M+K]+) and its respective dimmer adducts ([2M+NH4]+, [2M+Na]+, [2M+K]+). However, the observed m/z 343.1563 did not fit with the exact mass for the suspected protonated molecule (([M+H]+ = 343.1393). Then, it seems that the compound did not produce a stable protonated molecule. However, the ion at m/z 163.0762 could be explained by a neutral loss of hexose ($C_6H_{12}O_6$) from the suspected protonated molecule ($C_{16}H_{23}O_8$) and it was further fragmented to m/z 131.0502 ion (C_9H_7O) via neutral loss of methanol. Furthermore, HE

spectrum (see Fig. 5c) and MS/MS experiments (data not shown) corroborated that all the ions present in the spectrum came from the same compound. Additionally, it was concluded after MS/MS experiments that the in-source fragment at m/z 163 produced the m/z 131 ion at low collision energies, which in turns generated the m/z103 ion with a neutral loss of CO. Finally, the ion at m/z 343.1393 could be explained as a water adduct of the dimmer of the m/z 163.0762 fragment.
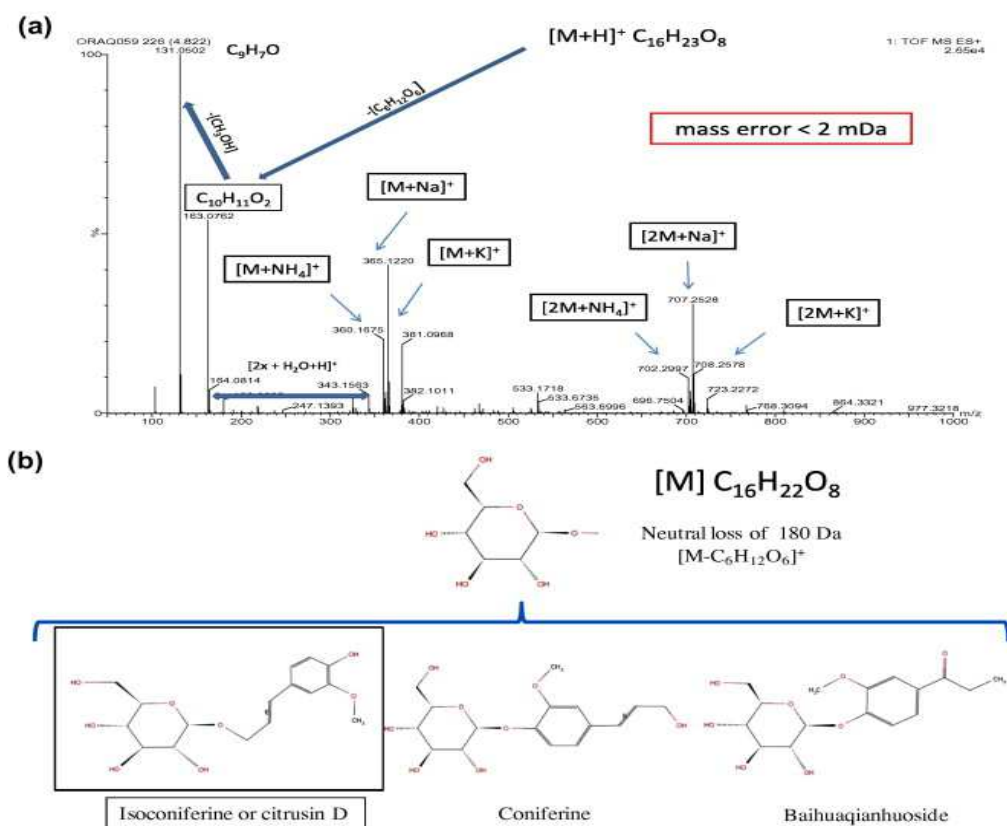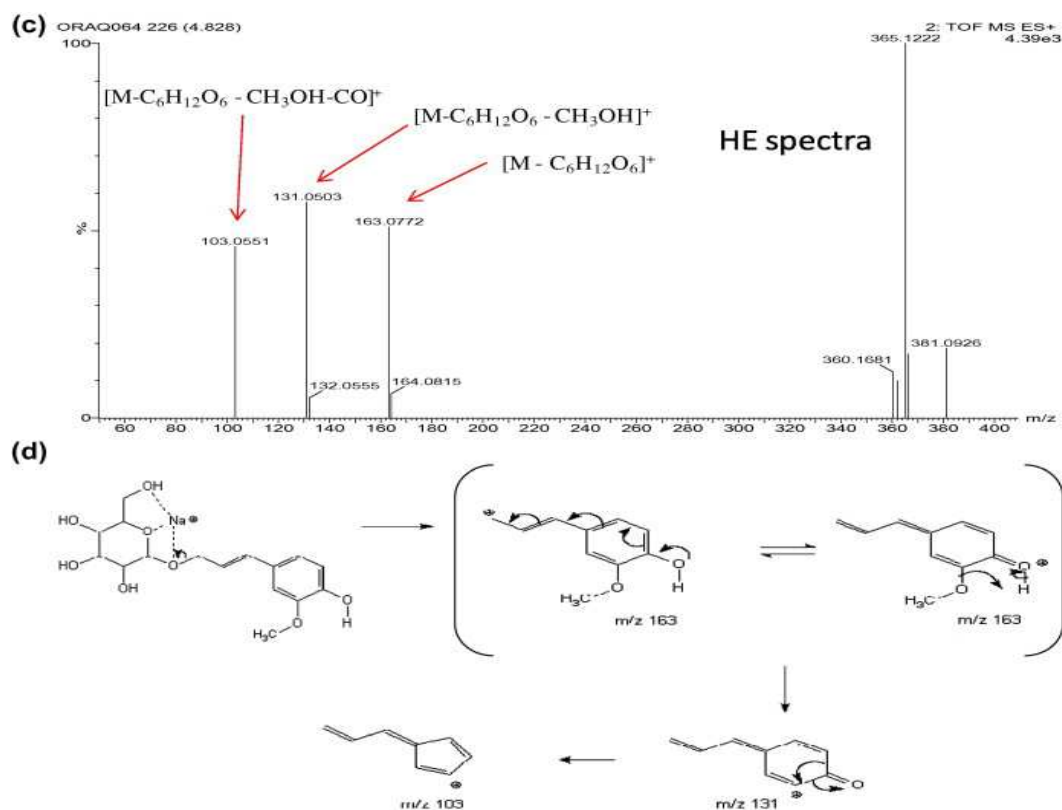
Fig. 5. (a) Spectrum of the feature/s of the candidate marker compound at RT 4.83 min, showing elemental composition and neutral losses/adducts of the most important ions and (b) candidate structures obtained for the molecular formula C16H22O8 in Reaxys database after substructure limitation as well as (c) HE spectra of the marker and (d) proposed fragmentation pathway for the compound selected (citrusin D).

The molecular formula found (C16H22O8) was searched in Reaxys database (Elsevier, NL) using a hexose (C6H12O6) substructure limitation and three results were obtained (see Fig. 5b). The most probable compound was citrusin D (or isoconiferine) as the observed neutral loss of 180 Da (C6H12O6) indicates an aliphatic attachment of the glycosyl group to the aglycone while the other two candidates, coniferine and baihuaqianhuoside, should shown the loss of 162 Da, common in aromatic linkages (Prasain, 2011). Furthermore, all fragment ions present in the HE spectrum could be easily explained for citrusin D (Fig. 5d). This compound has been reported to be present in "sweet oranges" peel, as the varieties analysed here (Matsubara, Sawabe, Ueno, Iizuka, & Okamoto, 2008). Unfortunately, to the best of our knowledge, there is no reference standard available for this compound and final confirmation could not be performed. However,

our data strongly support a tentative identification of citrusin D as the biomarker for differentiation oranges origin.

The fact that the main biomarker is present in the peel could be influenced by the environmental or the treatment applied to the oranges during harvesting. However, samples from two seasons and same post-harvest treatment were taken into account, minimising the environmental or manufacturing influence.

Other markers like M603T503 in Table 1a were also investigated as they were also representative of the sample differentiation although not so efficient marker as citrusin D according to the two seasons data. Only a tentative elemental composition was achieved and no well-fitting structure with the fragmentation data was found in the databases used (i.e., Reaxys, Chemspider and METLIN).

3.4. Negative ESI data

The same methodology was applied for the negative ionisation data (i.e. PLS-DA and OPLS-DA to easily obtain best markers). In this case, although separation was achieved in the PLS-DA model, markers relevance was slightly lower (data not shown), and then, data from negative ESI were finally discarded.

4. CONCLUSIONS

A metabolomic fingerprinting approach has been developed for the origin analysis of orange samples using an UHPLC-(Q)TOF MS system. Samples from Valencia and from the South hemisphere with similar varieties have been analysed to create a PLS-DA model that allowed the classification of 100% of samples. The model permitted to find useful marker/s to differentiate Valencia and foreign classes. The method was validated by increasing the number of samples (two seasons) and including Brazil origin in the foreign class. The most important marker was tentatively identified as citrusin D, which was elucidated by TOF MS data using MSE approach. Moreover, citrusin D was found as an optimal marker for two different years samples, while other potential markers showed a seasonal effect, corroborating its applicability independently on the collection time.

We suggest that this compound can be used to rapidly differentiate the Valencia oranges against southern ones by employing well-known techniques for target analysis, such as LC–MS/MS, more common in analytical laboratories, offering in this way a suitable approach to avoid fraud in this type of samples.

ACKNOWLEDGEMENTS

REFERENCES

Anderson and Smith, 2006
K.A. Anderson, B.W. Smith
Effect of season and variety on the differentiation of geographic growing origin of pistachios by stable isotope profiling

Journal of Agricultural and Food Chemistry, 54 (5) (2006), pp. 1747–1752

|

Arvanitoyannis et al., 1999
I.S. Arvanitoyannis, M.N. Katsota, E.P. Psarra, E.H. Soufleros, S. Kallithraka
Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics)

Trends in Food Science & Technology, 10 (10) (1999), pp. 321–336

Arvanitoyannis and Vaitsi, 2007
I.S. Arvanitoyannis, O.B. Vaitsi
A review on tomato authenticity: Quality control methods in conjunction with multivariate analysis (chemometrics)

Critical Reviews in Food Science and Nutrition, 47 (7) (2007), pp. 675–699

Cevallos-Cevallos et al., 2009
J.M. Cevallos-Cevallos, J.I. Reyes-De-Corcuera, E. Etxeberria, M.D. Danyluk, G.E. Rodrick
Metabolomic analysis in food science: A review

Trends in Food Science & Technology, 20 (11–12) (2009), pp. 557–566

Choi et al., 2010
M.Y. Choi, W. Choi, J.H. Park, J. Lim, S.W. Kwon
Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data

Food Chemistry, 121 (4) (2010), pp. 1260–1268

Consonni et al., 2009
R. Consonni, L.R. Cagliani, M. Stocchero, S. Porretta
Triple concentrated tomato paste: Discrimination between Italian and Chinese products

Journal of Agricultural and Food Chemistry, 57 (11) (2009), pp. 4506–4513

Cozzolino et al., 2003
D. Cozzolino, H.E. Smyth, M. Gishen
Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins

Journal of Agricultural and Food Chemistry, 51 (26) (2003), pp. 7703–7708

Cuadros-Inostroza et al., 2010
A. Cuadros-Inostroza, P. Giavalisco, J. Hummel, A. Eckardt, L. Willmitzer, H. Peña-Cortés
Discrimination of wine attributes by metabolome analysis

Analytical Chemistry, 82 (9) (2010), pp. 3573–3580

Dettmer et al., 2007
K. Dettmer, P.A. Aronov, B.D. Hammock
Mass spectrometry-based metabolomics

Mass Spectrometry Reviews, 26 (1) (2007), pp. 51–78

Díaz et al., 2011
R. Díaz, M. Ibáñez, J.V. Sancho, F. Hernández
Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry

Rapid Communications in Mass Spectrometry, 25 (2) (2011), pp. 355–369

Dunn and Ellis, 2005
W.B. Dunn, D.I. Ellis
Metabolomics: Current analytical platforms and methodologies

TrAC – Trends in Analytical Chemistry, 24 (4) (2005), pp. 285–294

Gómez-Ariza et al., 2005
J.L. Gómez-Ariza, M.J. Villegas-Portero, V. Bernal-Daza
Characterization and analysis of amino acids in orange juice by HPLC-MS/MS for authenticity assessment

Analytica Chimica Acta, 540 (1) (2005), pp. 221–230

Iijima et al., 2008
Y. Iijima, K. Suda, T. Suzuki, K. Aoki, D. Shibata
Metabolite profiling of chalcones and flavanones in tomato fruit

Journal of the Japanese Society for Horticultural Science, 77 (1) (2008), pp. 94–102

Lu et al., 2008
X. Lu, X. Zhao, C. Bai, C. Zhao, G. Lu, G. Xu
LC–MS-based metabonomics analysis

Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, 866 (1–2) (2008), pp. 64–76

Luykx and van Ruth, 2008
D.M.A.M. Luykx, S.M. van Ruth
An overview of analytical methods for determining the geographical origin of food products

Food Chemistry, 107 (2) (2008), pp. 897–911

Makris et al., 2006
D.P. Makris, S. Kallithraka, A. Mamalos
Differentiation of young red wines based on cultivar and geographical origin with application of chemometrics of principal polyphenolic constituents

Talanta, 70 (5) (2006), pp. 1143–1152

Malet, 2010
Malet, V. (2010). Dependence on 'old economy' mix fuels hangover in Valencia. *Financial Times*, March 3rd 2010.

Matsubara et al., 2008
Y. Matsubara, A. Sawabe, K. Ueno, Y. Iizuka, K. Okamoto
Structure and physiological activity of phenyl propanoid glycosides on kinkan (*Fortunella japonica*), sudachi (*Citrus sudachi*), yuzu (*Citrus junos* SIEB.), zabon (*Citrus gradis* OSBECK.) and unshiu (*Citrus unshiu*) peelings
Medical Journal of Kinki University, 1192 (2008), pp. 139–146

Moco et al., 2007
S. Moco, J. Vervoort, R.J. Bino, R.C.H. De Vos
Metabolomics technologies and metabolite identification

TrAC – Trends in Analytical Chemistry, 26 (9) (2007), pp. 855–866

Montero-Prado et al., 2013
P. Montero-Prado, K. Bentayeb, C. Nerín
Pattern recognition of peach cultivars (*Prunus persica* L.) from their volatile components
Food Chemistry, 138 (1) (2013), pp. 724–731

Nicholson et al., 2002
J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes
Metabonomics: A platform for studying drug toxicity and gene function

Nature Reviews Drug Discovery, 1 (2) (2002), pp. 153–161

Pendyala et al., 2007
G. Pendyala, E.J. Want, W. Webb, G. Siuzdak, H.S. Fox
Biomarkers for neuroAIDS: The widening scope of metabolomics

Journal of Neuroimmune Pharmacology, 2 (1) (2007), pp. 72–80

Perez et al., 2006
A.L. Perez, B.W. Smith, K.A. Anderson
Stable isotope and trace element profiling combined with classification models to differentiate geographic growing origin for three fruits: Effects of subregion and variety

Journal of Agricultural and Food Chemistry, 54 (13) (2006), pp. 4506–4516

Prasain, 2011
Prasain, J. (2011). *On-line information Ion fragmentation of small molecules in mass spectrometry*. <http://www.uab.edu/proteomics/pdf_files/2011/BGM744-01-19-11.pdf>.

Reid et al., 2006
L.M. Reid, C.P. O'Donnell, G. Downey
Recent technological advances for the determination of food authenticity

Trends in Food Science & Technology, 17 (7) (2006), pp. 344–353

Risticevic et al., 2008
S. Risticevic, E. Carasek, J. Pawliszyn
Headspace solid-phase microextraction-gas chromatographic–time-of-flight mass spectrometric methodology for geographical origin verification of coffee

Analytica Chimica Acta, 617 (1–2) (2008), pp. 72–84

Saavedra et al., 2000
L. Saavedra, A. García, C. Barbas
Development and validation of a capillary electrophoresis method for direct measurement of isocitric, citric, tartaric and malic acids as adulteration markers in orange juice

Journal of Chromatography A, 881 (1–2) (2000), pp. 395–401

Vaclavik et al., 2011
L. Vaclavik, A. Schreiber, O. Lacina, T. Cajka, J. Hajslova
Liquid chromatography–mass spectrometry-based metabolomics for authenticity assessment of fruit juices

Metabolomics (2011), pp. 1–11

4.2.1. Discusión de resultados

Este artículo refleja la eficacia del *metabolomics fingerprint* en el campo del control de calidad de alimentos. Concretamente, esta aproximación ofrece una herramienta para la correcta identificación de muestras de naranjas según su origen. Sin embargo, ésta es sólo una posible aplicación de una poderosa futura aproximación en el campo de la seguridad alimentaria.

A lo largo de esta tesis doctoral se ha aplicado XCMS para el pre-procesamiento de datos (Smith, Want et al. 2006). Este paquete de R (programa estadístico), además de ser totalmente gratuito, ha demostrado ser muy poderoso para el análisis de datos de MS, y más especialmente para LC-MS (Want and Metz 2010). Este paquete permite realizar los procesos de peak picking, agrupación y alineamiento de picos e integración de los mismos de una manera rápida, relativamente sencilla y, sobretodo, muy fiable (Smith, Want et al. 2006). Además, permite obtener un primer análisis univariante de los datos. Existen dos principales métodos de peak-picking dentro de XCMS, MatchedFilter y Centwave. Ambos se basan en la extracción secuencial de nw-XICs pero divergen en la detección de picos en éstos. Así pues, mientras MatchedFilter busca y ajusta el XIC a una curva Gaussiana, Centwave optimiza dicha detección basándose en el hecho de que los errores de masa en los diferentes espectros dentro de un pico cromatográfico son menores (o deben serlo) que aquellos aleatorios producidos por ruido químico o electrónico.

El uso del algoritmo Centwave es, en general, más favorable y robusto. Sin embargo, el correcto funcionamiento de éste depende precisamente de que el instrumento utilizado sea capaz de cumplir la condición, previamente mencionada, de reportar errores de masa muy bajos entre espectros de un mismo pico cromatográfico. En este trabajo se estudió la aplicabilidad de ambos algoritmos a los datos obtenidos, adquiridos mediante QTOF Premier (Waters). Los resultados reflejaron una mejor capacidad de detección del algoritmo MatchedFilter en base al número de picos encontrados y la veracidad de estos mediante inspección visual. El motivo de este hecho es precisamente el error de masa entre los diferentes espectros ya que este espectrómetro de masas está equipado con un detector tipo TDC, el cual es (o era)

muy sensible a la intensidad de la señal, mostrando errores de masa elevados para aquellos iones saturados o de muy baja intensidad. Precisamente la baja intensidad de la señal en los extremos de picos cromatográficos no excesivamente abundantes pero sin duda detectables implica que sean descartados por el algoritmo Centwave. En la figura 4.C. se especifican los parámetros utilizados en ambas metodologías además de los sucesivos procedimientos de agrupación, corrección del tiempo de retención e integración aplicados.

centwave method Peak picking

xs_n_1<-xcmsSet(method="centWave",peakwidth=c(5,20),snthresh=10,ppm=15)

xs_g1_n_1<-group(xs_n_1, bw=5)

xs_ret1_n_1<-retcor(xs_g1_n_1,f="s",p='m',missing=2)

xs_g2_n_1<-group(xs_ret1_n_1,bw=4)

xs_ret2_n_1<-retcor(xs_g2_n_1,f="s",p='m',missing=2)

xs_g3_n_1<-group(xs_ret2_n_1,bw=2)

xs_ret3_n_1<-retcor(xs_g3_n_1,f="s",p='m',missing=2)

xs_g4_n_1<-group(xs_ret3_n_1,bw=1)

xs_FP_n_1<-fillPeaks(xs_g4_n_1)

Matched Filter method Peak picking

xs_n_1_m<-xcmsSet(method="matchedFilter",snthresh=15,fwhm =15,step =0.02)

xs_g1_n_1_m<-group(xs_n_1_m, bw=30)

xs_ret1_n_1_m<-retcor(xs_g1_n_1_m,f="s",p='m',missing=2)

xs_g2_n_1_m<-group(xs_ret1_n_1_m,bw=5)

xs_ret2_n_1_m<-retcor(xs_g2_n_1_m,f="s",p='m',missing=2)

xs_g3_n_1_m<-group(xs_ret2_n_1_m,bw=2)

xs_ret3_n_1_m<-retcor(xs_g3_n_1_m,f="s",p='m',missing=2)

xs_g4_n_1_m<-group(xs_ret3_n_1_m,bw=1)

xs_FP_n_1_m<-fillPeaks(xs_g4_n_1_m)

Figura 4.C. Scripts aplicados en el pre-procesamiento de datos LC-(Q)TOF MS mediante el paquete XCMS de R.

Retomando el tema de la normalización de los datos, existen diversos modos de realizar dicha normalización. El primero de ellos, muy comúnmente aplicado, se basa en la utilización de uno o varios estándares internos. Al igual que ocurre durante el proceso de alineamiento de los tiempos retención, esta metodología presenta algunas desventajas que, especialmente en este tipo de aplicaciones, desaconsejan su uso. La primera desventaja es la asunción del equivalente comportamiento de todos los metabolitos, lo cual no es necesariamente cierto. El segundo inconveniente es la posible interferencia de dichos estándares internos con los metabolitos endógenos de la muestra por coelución, efecto matriz, etc. Este efecto puede no ser importante con el uso de un único (o muy pocos) compuesto de referencia pero de gran importancia si se pretende abarcar diferentes familias de metabolitos en todo el rango cromatográfico, y más especialmente en el acoplamiento LC-MS por su tendencia a sufrir efecto matriz.

Un método de normalización alternativo es el uso de algún/os compuestos endógenos. Este mecanismo es viable para muestras de orina, por ejemplo, ya que la propia muestra depende de la cantidad de orina excretada. Así pues, la efectividad de la corrección en base a la concentración de Creatina ha sido probada para la normalización de las muestras en base al diferente volumen de excreción y a las posibles derivas instrumentales. Sin embargo, este método no se antoja aplicable a muestras de alimentos.

Por último, la metodología evaluada, aunque finalmente no aplicada, en este artículo es la normalización en base a la mediana de todos los metabolitos detectados. Concretamente, mediante un algoritmo iterativo (Loess Normalization) se consigue igualar la mediana de la señal de todos los metabolitos en las diferentes muestras (Veselkov, Vingara et al. 2011). Obviamente, esta aproximación asume que el promedio de concentración de todos los metabolitos detectados en las diferentes clases es constante y, por tanto, que las diferencias entre ellas son prácticamente

despreciables. Aunque pueda parecer arriesgado, como se demuestra en el artículo, los cromatogramas de las diferentes muestras son indistinguibles y solo unas decenas de compuestos, de entre los miles de metabolitos detectados, presentan una diferencia significativa en las distintas clases. Por el contrario, no se distorsiona la muestra con compuestos exógenos y no se asume el idéntico comportamiento de los metabolitos ya que son los mismos endógenos los que se utilizan para la corrección.
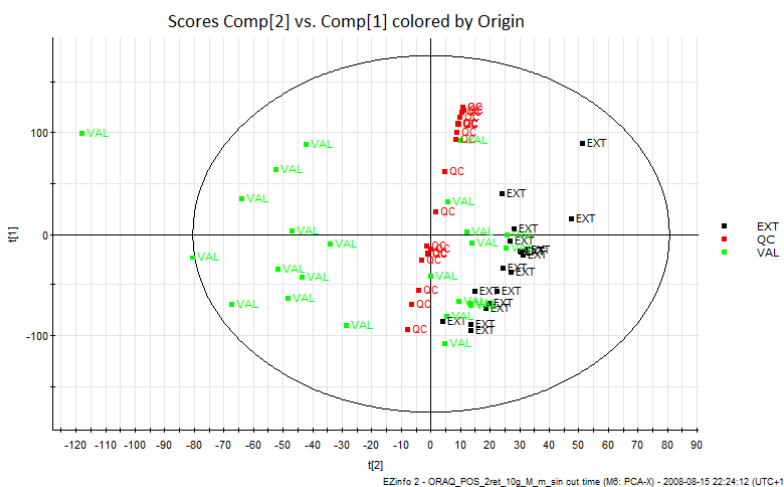
A) Normalized
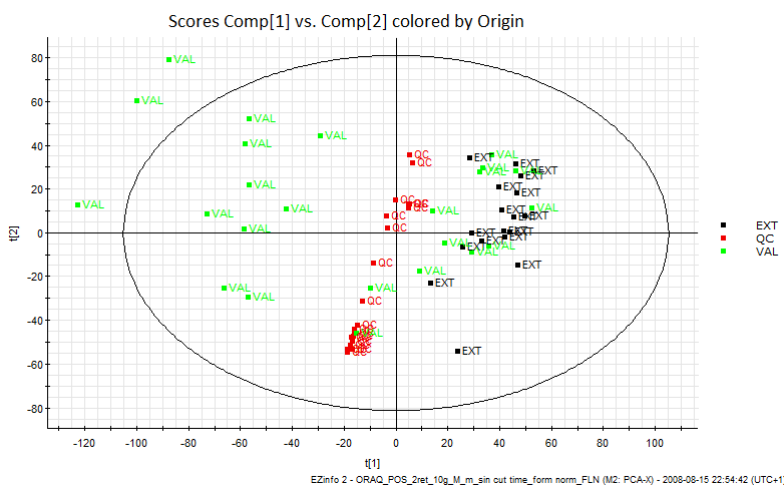
B) No normalized

Figura 4.D. Scores Plot (PC1 vs. PC2) del PCA de datos obtenidos del análisis de las muestras de naranja por LC-MS mediante ESI en modo positivo. En el Scores Plot (A) se muestran los datos normalizados mediante el algoritmo Loess Normalization mientras que en el (B) se modelan los datos crudos.

En este trabajo se comparó el efecto de la normalizacion de los datos obteniendo resultados similares en ambos casos, probablemente debido al relativamente bajo número de muestras analizadas por secuencia. En la figura 4.D. se muestran los PCA correspondientes a los datos crudos y normalizados, donde se puede observar que los resultados resultan muy similares.

Adicionalmente al proceso de normalización, el escalado de los datos (*Data Scaling*) suele ser necesario en el análisis multivariante de este tipo (datos MS). El proceso de escalado evita que picos de gran intensidad encubran picos de menor escala pero de igual importancia. Como se explicó en la introducción de este capítulo, el análisis multivariante (especialmente el PCA) reduce la dimensionalidad de los datos intentando explicar el máximo de varianza del conjunto. Por ende, aquellas variables que se encuentren a mayor escala tendrán una mayor varianza e influirán en mayor medida en la creación del modelo estadístico. Este efecto es sumamente común en datos de MS, el cual posee un relativamente alto rango dinámico y, por tanto, valores en rangos de más de 4 órdenes de magnitud pueden ser observados. Existen diferentes transformaciones que permiten minimizar este efecto. En este trabajo se probaron dos diferentes métodos: Varianza unitaria (*Unit Variance*) y Pareto. La aproximación Varianza Unitaria pretende igualar la varianza de todas las variables a 1, dando así el mismo peso a todas ellas en la creación del modelo aunque con la desventaja de hacerlo también para el ruido. El método Pareto proporciona una situación intermedia entre la transformación de Varianza Unitaria y los datos crudos al reducir la varianza a la desviación estándar de cada variable. En el caso de este trabajo no se apreció ninguna diferencia significativa entre los dos métodos de escalado pero si se observó una mejora en la clasificación de las muestras y los QCs respecto a los datos sin escalado.

Cabe destacar la importancia de la validación de la metodología, el modelo estadístico y los biomarcadores encontrados. En cuanto a la metodología, como ya se ha comentado previamente, la utilización de QCs, obtenidos como una mezcla de

todas las muestras analizadas, constituyen una herramienta muy eficaz para comprobar la eficiencia de la metodología y su reproducibilidad, así como para corregir y normalizar posibles efectos negativos.

En cuanto a la validación del modelo estadístico, y aunque no se detalla en el artículo, la metodología más comúnmente aceptada es la cross-validation. En la validación cruzada de K iteraciones (o K-fold cross-validation), los datos se dividen en K subconjuntos de muestras. Uno de los subconjuntos se utiliza como datos de prueba y el resto (K-1) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. Finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Este método es muy preciso dado que se evalúa, a partir de K combinaciones de datos de entrenamiento y de prueba, la capacidad predictiva del modelo sin renunciar a un determinado set de datos, permitiendo además evaluar la robustez del modelo. En la práctica, la elección del número de iteraciones depende de la medida del conjunto de datos. Lo más común es utilizar la validación cruzada de 10 a 20 iteraciones (10-fold cross-validation) lo cual implica que cada modelo generado en las diferentes iteraciones utiliza un 5-10% de las muestras como set de prueba. Existen otras metodologías aplicadas, como por ejemplo la utilización de un set de muestras para la creación del modelo y otro específico para la validación de éste, pero en este caso la limitación de muestras disponibles no permitió esta aproximación.

Por último, el error más común en este tipo de metodologías es la falta de validación de los marcadores encontrados, es decir, comprobar la eficacia del modelo mediante un set de muestras completamente diferente. En este caso, dicho procedimiento era extremadamente complicado debido al limitado número de muestras disponibles utilizadas para la creación del modelo estadístico y la baja representatividad de éstas al tratarse de la misma temporada. Así pues, se decidió utilizar un segundo set de muestras (correspondientes a la siguiente temporada de cultivo) para otorgar una mayor robustez y capacidad de predicción al modelo. Como resultado se obtuvo que algunos de los principales marcadores encontrados no resultaban serlo al combinar los datos de ambas temporadas. Por el contrario, el

marcador posteriormente identificado como Citrusin D fue corroborado con el set de datos global. Este resultado refleja la importancia de ir más allá en la investigación de los marcadores, la metodología y el modelo estadístico.

Se debe además comentar que la este marcador se propone como candidato para la discriminación de naranjas según su procedencia. Se requeriría un mayor número de muestras para poder afirmar que dicho metabolito es efectivamente un marcador inequívoco de la clasificación. Además, se debe aclarar que el objetivo no era encontrar un analito capaz de discernir el origen de la muestra, sino de dar una respuesta afirmativa o negativa a la procedencia de la Comunidad Valenciana. Por tanto, y debido a la enorme dificultad de conseguir muestras, el compuesto seleccionado, Citrusin D, puede únicamente ser denominado como candidato pero requeriría de un estudio más exhaustivo.

4.3. Artículo científico 6

"Told through the wine: an LC-MS interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics"

R. Díaz, H. Gallart-Ayala, J. V. Sancho, O. Nuñez, C.P.B. Martins, F. Hernández, S. Hernández-Cassou, J. Saurina and A. Checa.

Journal of Chromatography A, submitted.

Told through the wine: an LC-MS interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics

Ramon Díaz[a], Hector Gallart-Ayala[b], Juan V. Sancho[a], Oscar Nuñez[b], Claudia P.B. Martins[c], Félix Hernández[a], Santiago Hernández-Cassou[b], Javier Saurina[b] and Antonio Checa[b]

[a] Research Institute for Pesticides and Water, University Jaume I, Av. Sos Baynat S/N, 12071 Castellón, Spain

[b] Department of Analytical Chemistry, Universitat de Barcelona, Martí i Franquès 1-11, E-08028 Barcelona, Spain

[c] Thermo Fisher Scientific, 355 River Oaks Parkway, 95134, San Jose, CA, USA

ABSTRACT

This work focuses on the influence of the selected platform on the final annotated compounds in LC-HRMS non-targeted metabolomics. Differences between platforms included columns, mobile phases, gradients, chromatographs, mass spectrometers (Orbitrap and Q-TOF), data processing and marker selection protocol. A total of 42 wines samples from three different protected denomination of origin (PDO) were selected. Both platforms allowed the classification of samples according to their PDO at the feature level. At the annotated metabolite level, each platform proposed 9 (Orbitrap) and 8 (Q-TOF) annotated metabolites, identified by matching standards or MS/MS of the compound. None of the suggested metabolites was coincident between platforms. When screened on the raw data, 6 and 5 of these compounds were detected on the other platform with a similar trend. Some of the detected metabolites showed complimentary information when integrated on biological pathways. Through the use of some examples at the annotated metabolite level, possible explanations of this initial divergence of the results are presented. Despite the high focus of the present study, this work shows the complications that may arise on the comparison of non-targeted metabolomics platforms.

1. INTRODUCTION

Food characterization approaches are useful to increase our knowledge of products that regularly reach our tables. In recent years and thanks to the advances in technology and instrumentation, new powerful tools for the characterization and differentiation of foodstuff and beverages have emerged. Among them metabolomics, defined as the study of small molecules (<1500 Da) occurring in a given biological system has proven its utility in many research areas, including biology and physiology(Anas Kamleh, Spagou, J. Want 2011), plant science (Wolfender, Rudaz, Choi, Kim 2013) and food characterization and discrimination (Cevallos-Cevallos, Reyes-De-Corcuera, Etxeberria, Danyluk, Rodrick 2009; Checa, Saurina 2013). Particularly, several works have used metabolomics approaches in wine origin authentication (Serrano-Lourido, Saurina, Hernández-Cassou, Checa 2012; Springer, Riedl, Esslinger, Roth, Glomb, Fauhl-Hassek 2014; Versari, Laurie, Ricci, Laghi, Parpinello 2014).

Metabolomic methods can be classified into targeted and non-targeted modes. Targeted methods are mainly focused on a selected pre-defined group(s) of metabolites and can usually provide quantitative measurements of the selected compounds, while non-targeted metabolomics is an approach aiming at the pseudo-quantitative detection of many groups of compounds. Though absolute quantification is not possible, non-targeted methods are useful in the characterization (fingerprinting) of a system and classification of samples based on the differential abundance of the compound responses among the classes. These approaches have been applied to the identification and as authentication tools of complex alcoholic beverages, such as wine. Specifically, targeted approaches in wine characterization and discrimination have mostly been based on the analysis of wine enriched compounds such as organic acids(Gallagher, Heo, Lopez, Ray, Xiao, Umali, Zhang et al. 2012) and volatile(Fabani, Ravera, Wunderlin 2013) and phenolic compounds(Jaitz, Siegl, Eder, Rak, Abranko, Koellensperger, Hann 2010; Salvatore, Cocchi, Marchetti, Marini, de Juan 2013). In non-targeted approaches, on the other hand, the different families of metabolites can be simultaneously characterized and analyzed to provide a specific sample fingerprint with discriminating features that may(Arbulu, Sampedro, Gómez-Caballero, Goicolea, Barrio 2015; Diaz, Ibanez, Sancho, Hernandez 2011; Díaz, Pozo, Sancho, Hernández 2014; Ragone, Crupi, Piccinonna, Bergamini, Mazzone, Fanizzi, Schena et al. 2015) or may not(Springer, Riedl, Esslinger, Roth, Glomb, Fauhl-Hassek 2014) be later identified.

Though non-targeted metabolomics approaches are potentially much more informative than the targeted ones, in practice requires the annotation of the obtained features either by using databases or by matching with pure standards. As the metabolomic fingerprint is usually enough to discriminate the origin of the wines, further steps of identification are not usually carried out. This complicates the comparison with other laboratories and the translation of the method into other with a higher throughout or that are more cost-effective.

Some studies have been carried out in liquid chromatography-high resolution mass spectrometry (LC/HRMS) based metabolomics to study the influence of the instruments(Gika, Theodoridis, Earll, Snyder, Sumner, Wilson 2010; Glauser, Veyrat, Rochat, Wolfender, Turlings 2013) on the results, concluding that despite the expected differences, a high degree of overlap exist when the only difference present in the study is the mass spectrometer. However, differences between metabolomic platforms from different laboratories include several other factors, such as sample treatment, column, mobile phases and data processing. The influence of the analytical platform on the metabolic profiling level of urine samples has been recently evaluated (Martin, Maillot, Mazerolles, Verdu, Lyan, Migné, Defoort et al. 2015). High convergence between the spectral information produced by the different platforms was obtained, but an approach to evaluate how this translates into fully identified or annotated features was suggested (Martin, Maillot, Mazerolles, Verdu, Lyan, Migné, Defoort et al. 2015).

Thus in the present work we compared the performance of two different non-targeted LC-HRMS platforms from two different Spanish laboratories on the discrimination of Spanish red wines from three different protected denominations of origin (PDO). The ability of both platforms in the discovery of discriminating markers at the annotated metabolite level for the different PDO was inspected. Finally, potential discriminating metabolites individually suggested for each platform were compared and obtained results were put in a biological context.

2. MATERIALS AND METHODS

2.1 Reagents and chemicals

All reagents used were of analytical grade unless otherwise specified. Formic acid (99%) was obtained from Merck (Darmstadt, Germany) and Sharlau (Barcelona, Spain). LC-MS grade acetonitrile and water were purchased from Riedel-de-Haën (Seelze, Germany). HPLC grade methanol was purchased from Sharlau. Polyphenol standards were purchased from Sigma-Aldrich (St. Louis, MO). Mobile phases were filtered using 0.45 µm nylon membrane filters (Whatman, Clifton, NJ, US).

2.2 Samples and sample pretreatment

Forty-two commercial wines from three different PDO (14 Penedes, 14 Ribera del Duero and 14 Rioja) were purchased from retail stores. Samples were opened and divided into two different 5 mL aliquots, frozen at -20ºC and distributed between both platforms (**Figure 1**). Frozen samples were analyzed within a two weeks period after aliquotation. On the day of analysis, wines were thawed at room temperature and diluted 1:1 with double deionized water. Prior to injection, wine samples were filtered through 0.22 µm teflon membranes (Scharlab, Barcelona, Spain) for Platform#1 and centrifuged at 10000 *g* for Platform#2. Experimental workflow is summarized in **Figure 1**.
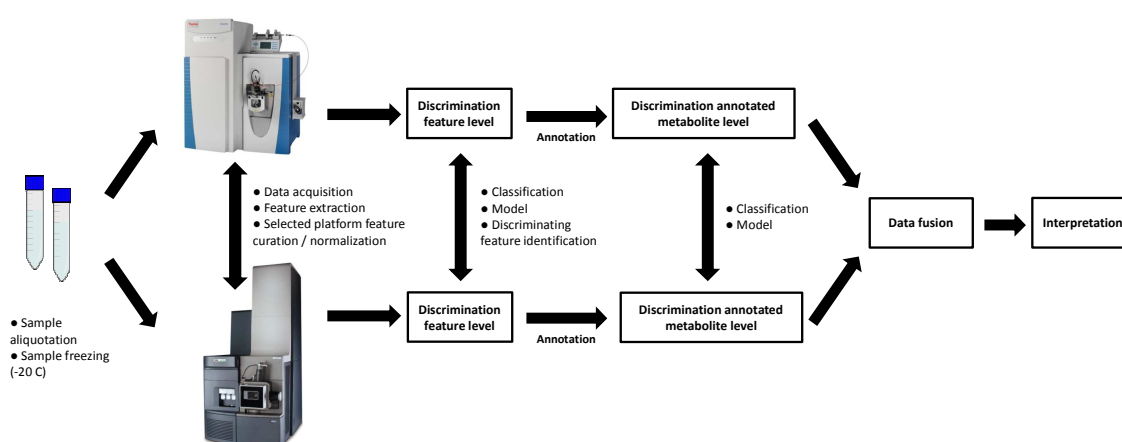

Fig. 1 Summary of the experimental workflow

2.3 UHPLC-HRMS methods

Platform#1. Orbitrap based metabolomics

An ultra-high performance liquid chromatography (UHPLC) system (Accela; Thermo Fisher Scientific, San José, CA, US) was used. The chromatographic separation was performed in a Hypersil Gold aQ (100 mm x 2.1 mm i.d., 1.9 µm particle size) column, from Thermo Fisher Scientific, using a gradient elution of 0.1% formic acid in water (solvent A) and acetonitrile (solvent B). The gradient elution started at 5% of solvent B (0 – 1 min isocratic step), followed by a linear gradient up to 25% B in 2.5 min and held during 1.5 min. Then, solvent B was increased up to 50% in 3 min. Finally another linear gradient up to 90% of solvent B was applied in 0.5 min and this

percentage was maintained during 1.5 min. The flow-rate was set at 600 µL min-1 and the column temperature and sample tray were held at 25°C and 4°C, respectively.

The Accela UHPLC system was coupled to a single-stage Orbitrap instrument (Exactive HCD; Thermo Fisher Scientific, Bremen, Germany) equipped with a heated electrospray (H-ESI II) ionization source. Nitrogen (purity > 99.999%) was used as sheath gas and auxiliary gas at flow-rates of 55 and 20 a.u. (arbitrary units), respectively. The ion transfer tube temperature was set at 270°C, the vaporizer temperature at 200ºC and the electrospray voltage was set at 3.5kV in negative mode. A scanning rate of 2 spectrum s-1 with a mass range of m/z 50 – 1000 with a mass resolving power of 50,000 Full Width Half Maximum (FWHM) (m/z 200) was used. In addition, for identification purpose selected wine samples were analyzed using "all ion fragmentation" (AIF) MS/MS acquisition mode applying collision energy of 28 eV in the high-energy collision dissociation cell (HCD).

Full instrument calibration was performed using a MSCAL5 ProteoMassT LTQ/FT-Hybrid ESI Pos/Neg (Sigma-Aldrich). External mass axis calibration without the use of the specific lock masses was employed.

For the accurate mass measurements, mass at the average of the chromatographic peak was obtained. The Xcalibur software version 2.2 (Thermo Fisher Scientific, San Jose, CA, US) was used to control the LC/MS system.

Platform#2. Q-TOF based metabolomics.

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (Xevo G2 Q-TOF, Waters Micromass, Manchester, UK), using an orthogonal Z-spray-ESI interface operating in negative ion mode. The UHPLC separation was performed using an Acquity UPLC BEH C18 column (100 × 2.1mm, 1.7 µm particle size) from Waters at a 300 µL/min flow rate. Mobile phase A and B consisted of water and methanol, both containing 0.01% formic acid. The percentage of organic modifier was changed as follows: 0 min, 5%; 12 min, 60%; 13.5 min, 90%; 14 min, 90%; 14.01 min, 5% in a total run time of 16 min. Nitrogen was used as the drying gas and nebulizing gas. The desolvation gas flow was set at 600 L/h. TOF-MS resolution was approximately 20000 at full width half maximum (FWHM) in V-mode at m/z 556.2771. MS data were acquired over an m/z range of 50–1200. A capillary voltage of 0.7 kV and a cone voltage of 25 V were used operating in negative ion mode Collision gas was argon 99.995% (Praxair, Valencia, Spain). The interface temperature was set to 350 °C and the source temperature to 120 °C. The column temperature was set to 40ºC.

For MS^E experiments, two acquisition functions with different collision energies were created. The first one, the low energy function (LE), selecting a collision energy of 4 eV, and the second one, the high energy (HE) function, with a collision energy ramp ranging from 15 to 40 eV in order to obtain a wider range of fragment ions. The LE and HE functions settings were for both a scan time of 0.2 s and an inter-scan delay of 0.05 s. The automated attenuated function was also selected to correct for possible peak saturations (extended mode).

Calibrations were conducted at a flow rate of 10 µL/min from *m/z* 50 to 1000 with a 1:1 mixture of 0.05 M NaOH: 5% formic acid diluted (1:25) with acetonitrile:water (80:20, v/v) and spiked with fungicide imazalil at a final concentration of 500 µg/L. For automated accurate mass measurement, the lock-spray probe was used, using as lock mass a solution of Leucine-enkephalin (2 µg/mL) in acetonitrile:water (50:50, v/v) at 0.1% formic acid pumped at 30 µL/min through the lock-spray needle. A cone voltage of 65 V was selected to obtain adequate signal intensity for this compound (~ 500 counts). The deprotonated molecule of Leucine-enkephalin, at *m/z* 554.2615 in negative mode, was used for recalibrating the mass axis and ensuring a robust accurate mass measurement along time. Identifications were carried out using MS^E acquisition mode(Diaz, Ibanez, Sancho, Hernandez 2011). MassLynx software was used for instrument control.

2.4 Data processing and statistical analysis

Platform#1. LC-HRMS/MS data was obtained in raw files format. For global analysis, raw files were first converted to mzXML format using MSConvert. Automatic peak detection and integration was performed using XCMS package for R. Centwave peak picking algorithm was used based on previous instrumental data analysis experience (peak width= 6-15 s, S/N threshold = 5, ppm = 15). Peak grouping was carried out using a density grouping method with an initial band width of 15 s. The resulting data matrix was then exported as comma separated values (CSV) files. Standard t-tests were conducted for univariate statistics using R (Version 3.1.0). SIMCA-P+® (Version 13, Umetrics AB, Sweden) software was used to perform multivariate statistical analysis. Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA) were applied to build descriptive and predictive models. The validity and robustness of the models were evaluated by $R^2(Y)$ and $Q^2(Y)$ parameters. Quality of cross-validation $Q^2(Y)$ value was assessed by using 2000 permutation tests as previously described(Westerhuis, Hoefsloot, Smit, Vis, Smilde, van Velzen, van Duijnhoven et al. 2008). Features in multivariate analysis were ranked according to their Variable Importance in Projection (VIP) values. Cross-validation models were used to calculate jackknife uncertainty of VIP.

Platform#2. UPLC-(Q)TOF MS data were converted from proprietary to netCDF format using Databridge application (within MassLynx v 4.1; Waters Corporation) and preprocessed using XCMS package for R. Centwave peak picking algorithm was used based on previous instrumental data analysis experience (peak width= 5-20 s, S/N threshold = 10, ppm = 25). Peak grouping was carried out using a density grouping method with an initial band width of 30 s. Then, peak retention time was aligned using a non-linear loess based correction fitting allowing a maximum of five missing samples to consider a peak group as "well behaved" and eligible to be included as an element of the retention time correction spine. Data was normalized to overcome instrumental drift due to cumulative interface contamination. Iterative Loess normalization was performed, making the median intensities of all features in every sample equal as previously described (Veselkov, Vingara, Masson, Robinette, Want, Li, Barton et al. 2011).

Multivariate analysis was carried out using SIMCA-P (Version 12, Umetrics, Sweden). Partial Least Square Discriminant Analysis (PLS-DA) was employed to distinguish samples according to their origin. Model was evaluated based on classification results after cross-validation. Then, orthogonal PLS-DA (OPLS-DA) analysis was used to highlight the markers to differentiate between classes.

2.5. Variable annotation

For simplification purposes only compounds of plant origin were kept for both platforms. When possible, compounds were confirmed by matching retention time and spectra of analytical standards. Otherwise characteristic fragments in MS/MS were screened for each compound.

Platform#1. Features showing significant VIP values combined with low jackknife errors (VIP-jackniffe > 1) were submitted to annotation process.

Platform#2. For candidate markers selection OPLS-DA S-Plot was used confronting every group with the rest. Those features with p(corr) > 0.7.

For both platforms, the combination of exact mass of the precursor ions obtained in full scan mode in combination with the HRMS(/MS) fragmentation was used for identification purposes. METLIN (http://metlin.scripps.edu) and the human metabolome database (http://www.hmdb.ca), this last one just focused on metabolites from food source, were used as databases for annotation.

3. RESULTS AND DISCUSSION

This section includes a summary of the performance of both platforms separately on the discrimination of wine samples according to its PDO as well as on the identification of the compounds driving the separation between groups.

3.1 Feature level discrimination

Initially, the discrimination abilities of both platforms using all extracted features were inspected. As it can be seen in the plot of scores (**Figure 2a-b**), separation between PDO was achieved for both experimental setups. Normalization used in Platform#2 resulted in a higher control of the experimental variation as shown by the clustering of the QCs. In both platforms, separation of the three classes was observed across PC1 for Penedes wines with a higher degree of overlap between Ribera del Duero and Rioja. Platform#2 provided better discrimination between classes.

Classification of samples into PDO was accomplished by means of PLS-DA and OPLS-DA for Platforms 1 and 2 respectively. In both cases, cross-validated models using all extracted features allowed the correct classification of the wines according to PDO (**Table 1**). Presence of overfitting due to the high feature to sample ratio was tested by comparing the obtained Q^2 to the values of 2000 permutated models(Westerhuis, Hoefsloot, Smit, Vis, Smilde, van Velzen, van Duijnhoven et al. 2008). The value of Q^2 obtained was lower than the value of the selected model in all cases (**Table 1**) suggesting no overfitting (p < 1/2000). This is also supported by the plot of the distribution of the permutated models in all cases (Supplementary **Figure 1**).

Table 1. Summary of PLS-DA models obtained for Platform 1 (Q-Exactive) and 2 (Q-TOF) working in the different chromatographic/MS models

		Penedes				Ribera				Rioja		
Platform	LV	$Q^{2\ a}$	p^b	%CC^c	LV	$Q^{2\ a}$	p^b	%CC^c	LV	$Q^{2\ a}$	p^b	%CC^c
1	1+1	0.89	<5x10^{-4}	100	1+1	0.83	<5x10^{-4}	100	1+1	0.72	<5x10^{-4}	100
2	1+0	0.86	<5x10^{-4}	100	1+1	0.86	<5x10^{-4}	100	1+1	0.77	<5x10^{-4}	100

a) Q2 obtained by the sevenfold cross-validation procedure obtained in SIMCA
b) Probability to obtained a Q^2 value equal or above to the one obtained for a specific model model when 2000 permutations are performed
c) Percentage of correctly classified samples

3.2. Annotated metabolite level discrimination

The top discriminating features between PDO classes were annotated using the selected protocol for each platform (See Section 2.5). Annotation was performed both after operator visual spectral inspection and using automated procedures based on accurate mass, referencing to public or in-house databases and, when possible, using commercially available compounds. Annotated compounds that were not from plant origin or presented an improbable retention time (*e.g.* highly retained polar compounds) were excluded. A total of 9 and 8 features were identified for Platform 1 and 2 respectively (**Table 2**). None of the annotated features was common between platforms.

PCA models built using only annotated features on each platform resulted in a clear class separation for Platform 2 (**Figure 2d**). In the case of Platform#1, separation between Ribera del Duero and Rioja was not complete using the unsupervised model (**Figure 2c**). OPLS-DA models built and validated as described in Section 3.2 using annotated metabolites for each platform provided similar results to the ones obtained when all features were used (**Table 2**).
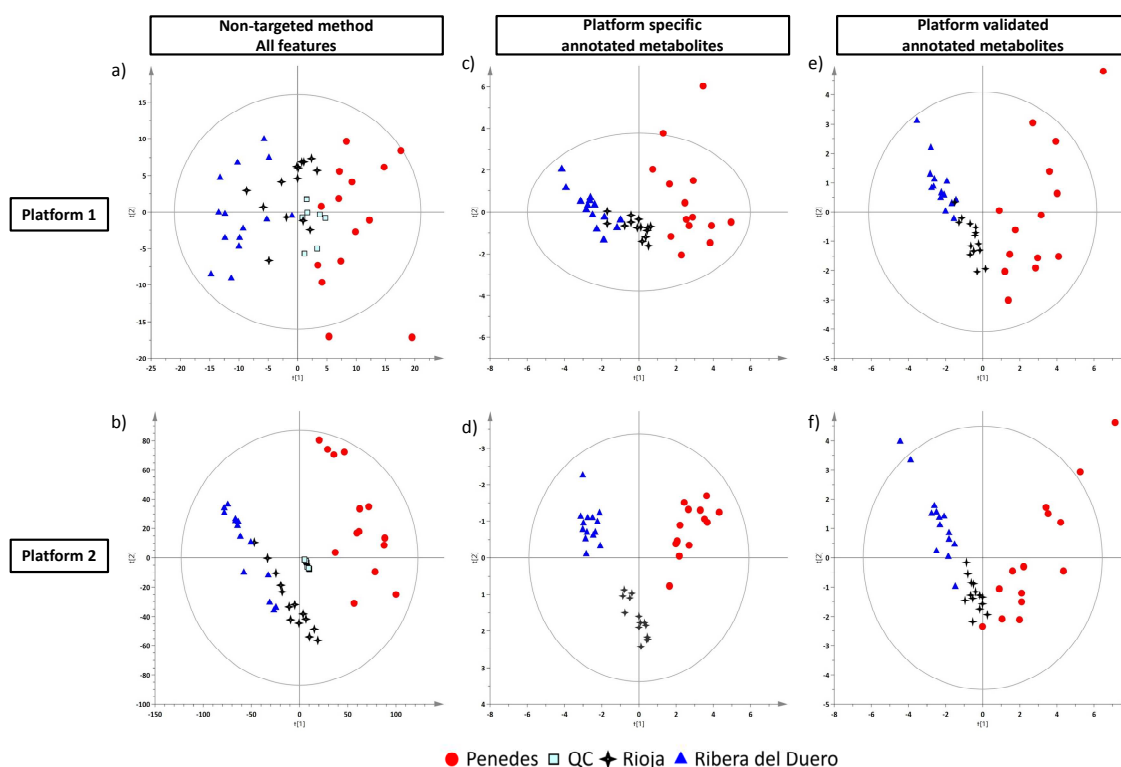

Fig. 2 Plots of scores obtained for the PCA analysis for Platforms 1 and 2 at the all features level (a-b), platform specific annotated metabolite level (c-d) and interplatform validated metabolites (e-f). In all cases, samples are coloured by PDO.

3.3. Interplatform annotated metabolites discrimination

Last part of the workflow consisted of the comparison of the annotated discriminating features between Platforms 1 and 2. Estimated retention times were obtained by injecting a mixture of polyphenols in both platforms (**Figure 3a**). Annotated compounds for each of the platforms presenting mass errors below 5 ppm at the expected retention time +- 15% were screened in the raw files of the other platform and areas were extracted (**Figure 3a**). Compounds annotated in one platform that were validated in the other are presented on **Table 2**. A total of 7 and 6 features from Platform 1 and 2 were identified on Platform 2 and 1 respectively. The resulting PCA model performed for annotated values of the annotated features common for both platforms showed a high degree of similarity between them (**Figure 2e-f**). OPLS-DA (**Table 1**) models performed for each platform only using features identified in both platforms also allowed the correct classification of all samples according to PDO.

Platform#1 found increased levels of the flavanols catechin and epicatechin in Penedes wines (**Figure 4 a-b**) and of their derived compounds gallocatechin and epigallocatechin in Ribera del Duero wines (**Figure 4 c-d**). Both findings could be reproduced in Platform#2. A digalactoside of myricetin was found increased in Ribera del Duero wines (**Figure 4e**) by Platform#2. Integration of the compound in Platform#1 resulted in an equal significant change, but fold change relative to the other two PDO was 12 times lower (**Figure 4e**).

3.4 Discussion

The present study aimed at the comparison of the performance of two different metabolomics platforms on the identification of discriminating compounds in wine authentication. Samples belonging to three different PDO that had already been discriminated according to their (poly)phenolic fingerprint were selected in order to address two questions. The main objective was to compare the impact of the metabolomic workflow on the discriminating compounds reported. Then we aimed to inspect how an LC-MS based metabolomics approach could improve and complement the generated knowledge of a less specific LC-UV signal at the metabolite level. Comparisons of the results show that despite the ability of platforms to discriminate between samples both at the spectral and at the annotated metabolite level, a strong divergence between the annotated metabolites involved in the discrimination can occur. Interestingly, most of these discriminating compounds could be detected on the complimentary platform when they were screened in a targeted manner.

Table 2. Metabolites separately identified in platform 1 and 2 indicating where it is validated (i.e. identified) in the other platform. FC = Fold-change.

| RT (s)	m/z (Theoric)	Δm (ppm)	Compound	P.D.O	FC (p-val)	Validated? (FC / p-val)
			Platform 1			
199	289.0720 (289.07176)	1.4	Epicathechin	P	2.06 (<0.0001)	YES (1.94/<0.0001)
82	173.0449 (173.0455)	-3.5	Shikimic acid	P	3.02 (<0.0001)	YES (6.45/9.40E-03)
49	159.0292 (159.0299)	-4.4	2-Methyl-4-oxopentanedioic acid	P	1.70 (<0.0001)	NO
48	169.0135 (169.0142)	-4.1	Gallic acid	P	1.54 (<0.0001)	NO
85	305.0668 (305.0667)	0.3	Gallocatechin	RD	1.72 (<0.0001)	YES (2.03/<0.0001)
161	305.0668 (305.0667)	0.3	Epigallocatechin	RD	1.75 (<0.0001)	YES (2.13/<0.0001)
172	329.0880 (329.0878)	0.6	3'-Glucosyl-2',4',6'-trihydroxyacetophenone	RD	1.71 (<0.0001)	YES (1.30/<0.0001)
240	389.1238 (389.1242)	-1.0	trans-piceid acid	P	3.87 (<0.0001)	YES (3.10/<0.0001)
273	389.1245 (389.1242)	0	cis-piceid acid	P	4.18 (0.0008)	NO
			Platform 2			
562	287.0927 (287.0925)	0.6	6'-Hydroxyangolensin or 1-(3-Methoxy-4-hydroxy)-phenyl-6,7-dihydroxy	P	4.78 (<0.0001)	YES (62.38/<0.0001)
276	481.0966 (481.0988)	-4.5	(-)-Epigallocatechin 3' (or 7)-glucuronide	RD	7.99 (<0.0001)	NO
219	165.0186 (165.0193)	-4.3	3,4-Methylenedioxybenzoic acid or Terephthalic acid	RIO	3.04 (<0.0001)	NO
618	403.1386 (403.1398)	-3.2	4'-Methylresveratrol 3-glucoside	P	32.56 (<0.0001)	YES (22.95/<0.0001)
551	431.1335 (431.1348)	-3.0	Trichocarposide	P	10.30 (<0.0001)	YES (19.40/<0.0001)
581	535.1816 (535.1821)	-1.0	8-Hydroxypinoresinol glucoside	P	17.28 (<0.0001)	YES (3.60/0.0005)
339	641.1349 (641.1359)	-1.7	Myricetin 3,3'-digalactoside	RD	30.12 (<0.0001)	YES (2.24/<0.0001)
325	655.1147 (655.1152)	-0.7	Gossypetin 8-glucuronide 3-glucoside	RD	26.91 (<0.0001)	NO
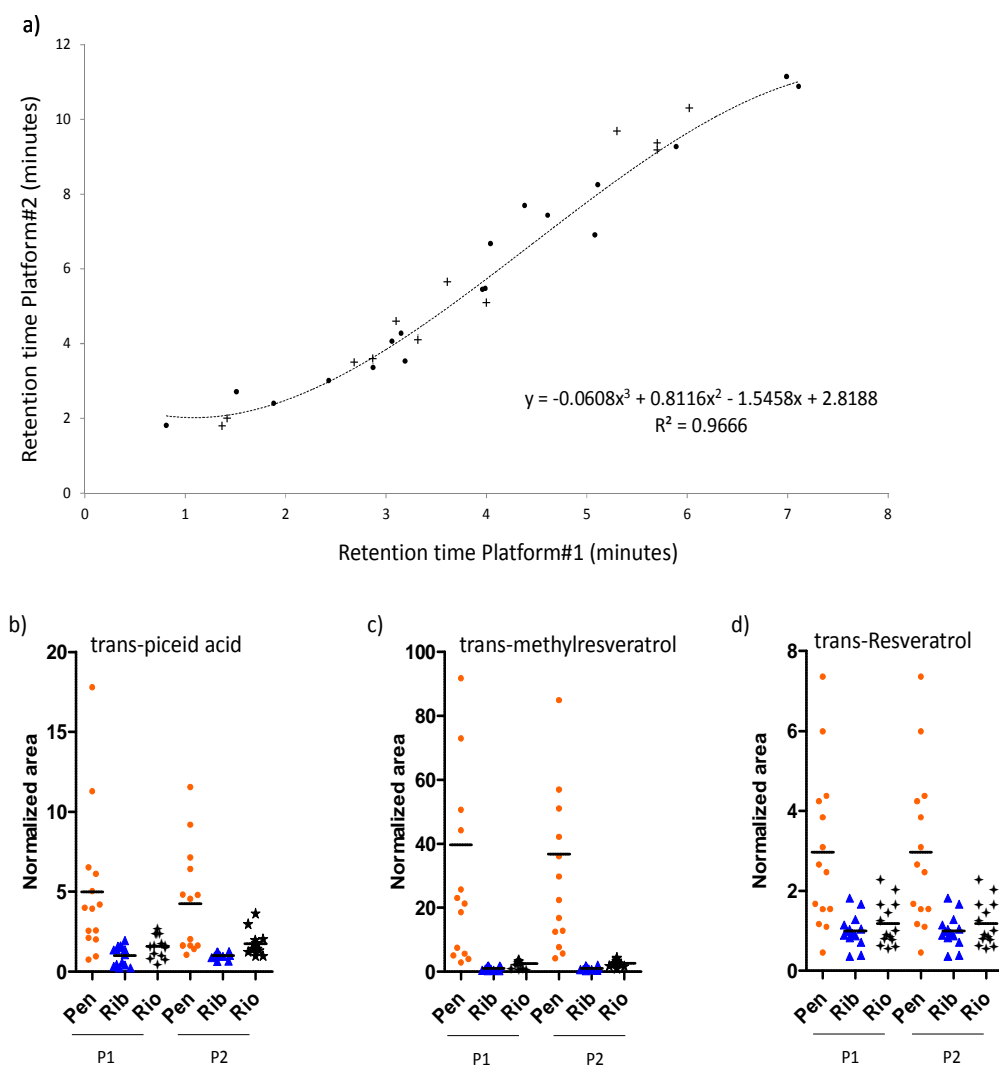
Fig. 3 (a) Retention time correlation between both platforms. Function used for retention time estimation of the markers found in the alternative platform. The list of reference standards and RT used for this function is shown in Table S1. (b-d) Extracted areas of the different resveratrol derived markers as a function of their PDO. To perform the comparison, all areas were normalized to the mean of the lower abundant class for each platform.

For Penedes wines, one resveratrol derived molecule was identified by each platform and validated by the other (**Table 2, Figure 3b-c**). Though it was not originally identified as a

discriminating compound by neither of the platforms, trans-resveratrol was also found elevated in Penedes wines by both platforms (**Figure 3d**).

Samples were analysed using two different HRMS instrument configurations widely employed in metabolomics analyses, Orbitrap (Platform#1) and Q-TOF (Platform#2). Reversed phase chromatography was selected based on the nature of compounds that were expected drive the separation, mainly polyphenols, though specific columns were used by each platform. For the same reason, negative mode was selected as the ionization mode.

First comparison was performed on the extracted metabolites. Both platforms used a QC pooled from all samples to control for reproducibility within batches. On Platform#1, no algorithm was employed for data correction and only compounds presenting a %CV≤30 on the QCs were submitted for later analysis. On Platform#2, compounds were corrected. Data were normalized by the locally weighted scatter plot smoothing (LOESS) normalization method which adjusts the local median of log fold changes of peak intensities between samples in the data set to be approximately zero across the whole peak intensity (Veselkov, Vingara, Masson, Robinette, Want, Li, Barton et al. 2011). This expectedly resulted in a lower number of extracted features for Platform#1, though results before curation were more similar. In addition, m/z obtained from Platform#1 were proportionally lower than the ones obtained from Platform#2. This observation is in accordance with previous experiments where an Orbitrap and Q-TOF platforms were used (Dervilly-Pinel, Weigel, Lommen, Chereau, Rambaud, Essers, Antignac et al. 2011).

In order to select the discriminating compounds, the ability of both platforms to discriminate between PDO on the feature level was inspected. In both platforms, Penedes samples could be separated from the other two PDO across PC1 using unsupervised PCA. This is in accordance with our previous observation based only in the LC-UV profiles of the samples (Serrano-Lourido, Saurina, Hernández-Cassou, Checa 2012). Expectedly, normalization strategy used on Platform#2 resulted in a better clustering of the QC (**Figure 2b**). This higher degree of clustering was also observed for samples belonging to a PDO. Ribera del Duero and Rioja samples could not be completely separated by any of the platforms by means of PCA. This is probably a result of the higher similarities in climatological conditions resulting from a closer geographical location of the PDO and also from the homogeneity on the grape variety. However, Penedes samples obtained from Tempranillo grape variety did not cluster with the other two PDO which shows that growing conditions do affect more the fingerprint than grape type in terms of PCA variance. As expected from the obtained PCA separation, the supervised OPLS-DA methodology allowed classifying all samples in the cross-validated model for both platforms (**Table 1**). Despite the large variable to sample ratio, especially on Platform#2, no signs

of overfitting were suggested when obtained models were compared to others built with permutated classed (Supplementary **Figure 1**).

Next, each platform submitted its top selected discriminating features to annotation. Because of the high number of potential discriminating compounds obtained on both cases, only metabolites from plant origin were kept. In addition, when more than three possible metabolites were returned by the database, the feature was also discarded. Though by using this approach most potentially discriminating features were excluded (See **Table 2**), this allowed to focus on the proper identification of the compounds by means of MS/MS. It also facilitated the later inter-platform comparison and interpretation of the results in biological pathways that is the desirable endpoint of most metabolomic studies. None of the annotated features for each platform was replicated on the other.

Though classification using annotated features was not the aim of the present work, models built including only annotated features also allowed the discrimination between classes (**Table 1, Figure 2c-d**). This shows that though the restrictions imposed in the annotation drastically reduced the number of variables to perform the model, annotated metabolites were still important for the performance of the separation in both platforms. None of the platforms could identify characteristic compounds of Rioja wines based on the selected criteria. Separation for this class on the metabolite level was thus based on the lack of the characteristic annotated compounds for each of the other two PDO.

The final goal of the workflow consisted of the integration of the results from both platforms. In order to facilitate this, a series of polyphenol standards covering the whole chromatographic range were injected and used to align compounds between the platforms based on their retention times (**Figure 3a**). Compounds that were not originally identified on each platform were originally searched in the extracted tables and later on the raw chromatogram files. Though none of the compounds independently identified for each platform was coincident, 6 of 9 compounds from Platform#1 could be identified on Platform#2 presenting the same trend when inspected on a targeted manner. In the case of Platform#2, 5 of 8 compounds could be identified on Platform#1. These results show the influence of the platform, including the whole pipeline of analysis on the final result of the metabolomics workflow. The aforementioned capability of each platform to detect lower or higher masses was also reflected on the distribution of the masses for identified compounds. Comparison of the plots of scores obtained from both platforms (**Figure 2 e-f**) highlight the high coherence between the obtained results at the annotated metabolite level. This consistency is also maintained for the individual metabolites

(**Figures 3 b-c and 4**). Possible reasons for the lack of reproducibility for some of the identified metabolites as well as the interpretation of the results are discussed below.

Each platform was able to pick one resveratrol derived compound as discriminating feature for Penedes wines (**Figure 3 b-c, Table** 2). In Platform#1 both cis and trans-Piceid acids (both glucosides of resveratrol) were identified. This is in accordance with our previous publication (Serrano-Lourido, Saurina, Hernández-Cassou, Checa 2012)), where trans- and cis-Piceid acids were also identified. Interestingly, in Platform#2, only transPpiceid acid could be validated. On the other hand, Platform#2 identified a methylated form of the glucoside that presented a much higher fold change. The compound was annotated using the accurate mass of the deprotonated molecule and the presence of a methyl-resveratrol fragment (**Supplementary Figure 2a**). As two different compounds presented the same accurate mass and fragment, trans- form was selected as the annotated compound based on the retention time pattern of both resveratrol and Piceid acid, as they elute before the cis one (**Figure 3a**). An inspection of the compound on the raw data of Platform#1 showed that the elevated levels of methyl resveratrol glucoside could also be found using this platform. Further inspection of the data table showed that the compound was discarded before submitting to the multivariate analysis because it presented a high %CV on the QCs and thus it was not originally identified. This example shows one of the risks of excluding features by a fixed %CV value on the QCs. Compounds presenting high fold changes can still be identified despite presenting high %CV. As it was available on the original polyphenol mixture, trans-resveratrol was also screened on the samples (**Figure 3d**). Though none of the platforms identified it as a discriminating compound, pattern shown in Penedes wines strengthens the results found for its derivate compounds. Low absorbance of resveratrol was the most likely reason why the compound was not identified using the LC-UV profile. In the case of LC-HRMS platforms, a possible explanation is the correlation with the other resveratrol derived compounds that showed stronger differences.

Independent flavanol related compounds were also identified by both platforms (**Figure 4**). Catechin related flavanols where identified in Platform#1 and later validated on Platform#2. While Catechin and Epicathechin were elevated in Penedes wines, their hydroxylated forms in the 3´positions (Gallo- and Epigallocatechin) were increased in Ribera del Duero wines. On the other hand, on Platform#2, elevated levels of a Myricetin digalactoside derived compound, were detected in Ribera del Duero wines. In our previous study, Myricetin and one of its glucosides were found increased in Ribera del Duero wines. As in the case of the resveratrol derived compounds, the use of the non-targeted platform reinforces the idea of an upregulation of this part of the pathway (**Figure 4**) on this PDO. Again, though Myricetin was not originally

identified by none of the platforms, manual integration of the compound showed its increased levels in Ribera del Duero wines. Interestingly in this case, fold changes found by Platform#2 were 10 times higher relative to Platform#1. One of the possibilities is the presence of interfering coeluting compounds causing ion suppression and affecting robust quantification of this compound in Platform#1 (**Supplementary Figure 2b**). This shows an important issue that may arise when comparing non-targeted, but also targeted metabolomics platforms where different chromatographic methods are used. In the present case, this strengthens the confidence of the found markers as they were obtained with two different chromatographic methods. On the other hand, when results are not validated, it may be difficult to make a decision on which result is the correct one if a labelled internal standard for the compound is not included on the analysis. Sample pre-treatment is also expected to introduce ion suppression related differences, though in the present case the different treatments carried out (filtration vs centrifugation, See Section 2.2) were not expected to introduce much differences.

Another derivate from (epi)gallocatechin, a glucuronide was originally proposed as one of the candidate metabolites by Platform#2 (**Table 2**) that could not be validated by Platform#1. Thus, no further identification was attempted.

In the case of Gallic acid, which was also identified as a marker for Penedes wines (Serrano-Lourido, Saurina, Hernández-Cassou, Checa 2012), Platform#1 was able to confirm elevated values on this PDO while on Platform#2 these increases could not be validated as the compound, included on the polyphenol mixture, was not found at the expected retention time. The closeness of the compound to the void volume, together with its relative low mass is a possible cause of this lack of detection.

4. CONCLUSIONS

Final identification of the metabolites remains one of the bottlenecks of non-targeted metabolomics platforms. Through the use of a restricted system were only a family of compounds as a fraction of the metabolome was used to facilitate the comparison, this work shows the difficulties of obtaining robust results in terms of annotated metabolites at the end of the workflow. These complications are expected to increase when carrying out a higher coverage of the metabolome. On the other hand, we show that the use of different platforms allows both complement and validate the results as compared to single platform based approaches.
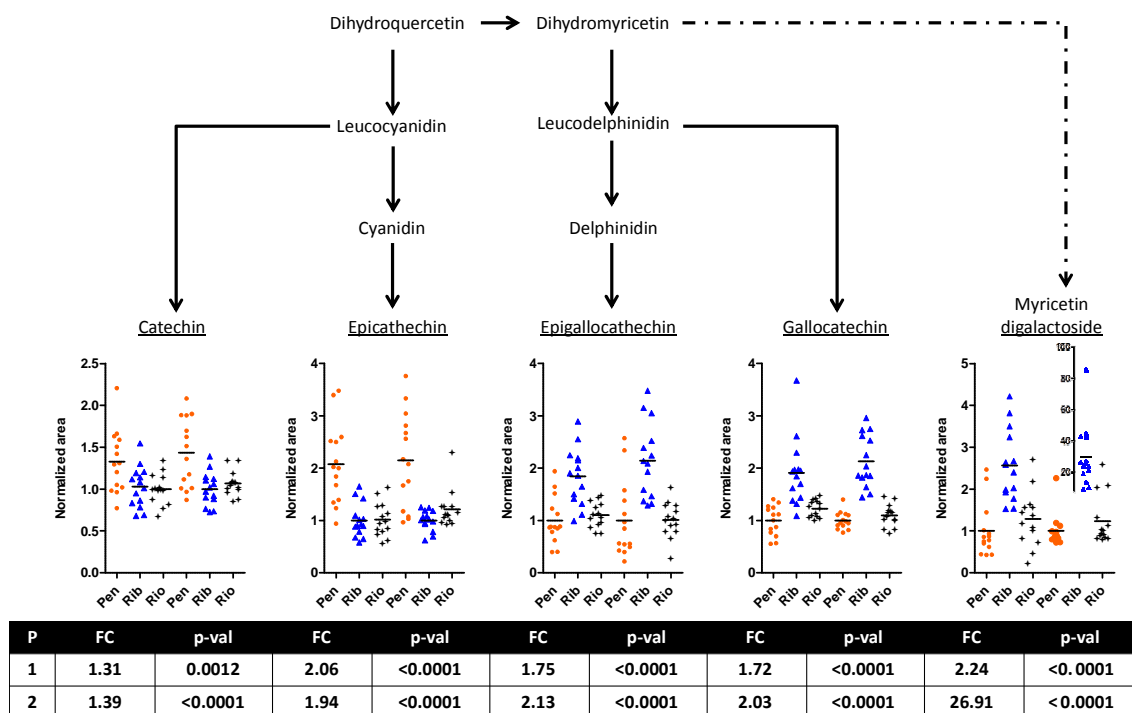
P	FC	p-val	FC	p-val	FC	p-val	FC	p-val	FC	p-val
1	1.31	0.0012	2.06	<0.0001	1.75	<0.0001	1.72	<0.0001	2.24	<0.0001
2	1.39	<0.0001	1.94	<0.0001	2.13	<0.0001	2.03	<0.0001	26.91	<0.0001

Fig. 4 Extracted areas of the catechin related compounds found increased in Penedes (a-b) and Ribera del Duero (c-e) wines. To perform the comparison, all areas were normalized to the mean of the lower abundant class for each platform.

5. REFERENCES

Anas Kamleh, M., K. Spagou, E. J. Want (2011). Metabolic Profiling in Disease Diagnosis, Toxicology and Personalized Healthcare. Current Pharmaceutical Biotechnology 12, 976-995 doi:10.2174/138920111795909069

Arbulu, M., M. C. Sampedro, A. Gómez-Caballero, M. A. Goicolea, R. J. Barrio (2015). Untargeted metabolomic analysis using liquid chromatography quadrupole time-of-flight mass spectrometry for non-volatile profiling of wines. Analytica Chimica Acta 858, 32-41 doi:http://dx.doi.org/10.1016/j.aca.2014.12.028

Cevallos-Cevallos, J. M., J. I. Reyes-De-Corcuera, E. Etxeberria, M. D. Danyluk, G. E. Rodrick (2009). Metabolomic analysis in food science: a review. Trends in Food Science & Technology 20, 557-566 doi:http://dx.doi.org/10.1016/j.tifs.2009.07.002

Checa, A., J. Saurina (2013). Chapter 6 - Metabolomics and PDO. In: Miguel de la, G., G. Ana (eds) Comprehensive Analytical Chemistry. vol Volume 60.(pp123-143). Elsevier.

Dervilly-Pinel, G., et al. (2011). Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves. Analytica Chimica Acta 700, 144-154 doi:http://dx.doi.org/10.1016/j.aca.2011.02.008

Diaz, R., M. Ibanez, J. V. Sancho, F. Hernandez (2011). Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 25, 355-69 doi:10.1002/rcm.4860

Díaz, R., O. J. Pozo, J. V. Sancho, F. Hernández (2014). Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. Food Chemistry 157, 84-93 doi:http://dx.doi.org/10.1016/j.foodchem.2014.02.009

Fabani, M. P., M. J. A. Ravera, D. A. Wunderlin (2013). Markers of typical red wine varieties from the Valley of Tulum (San Juan-Argentina) based on VOCs profile and chemometrics. Food Chemistry 141, 1055-1062 doi:http://dx.doi.org/10.1016/j.foodchem.2013.04.046

Gallagher, L. T., et al. (2012). Pattern-based discrimination of organic acids and red wine varietals by arrays of synthetic receptors. Supramolecular Chemistry 24, 143-148 doi:10.1080/10610278.2011.638379

Gika, H. G., G. A. Theodoridis, M. Earll, R. W. Snyder, S. J. Sumner, I. D. Wilson (2010). Does the Mass Spectrometer Define the Marker? A Comparison of Global Metabolite Profiling Data Generated Simultaneously via UPLC-MS on Two Different Mass Spectrometers. Analytical Chemistry 82, 8226-8234 doi:10.1021/ac1016612

Glauser, G., N. Veyrat, B. Rochat, J.-L. Wolfender, T. C. J. Turlings (2013). Ultra-high pressure liquid chromatography–mass spectrometry for plant metabolomics: A systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. Journal of Chromatography A 1292, 151-159 doi:http://dx.doi.org/10.1016/j.chroma.2012.12.009

Jaitz, L., et al. (2010). LC–MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage. Food Chemistry 122, 366-372 doi:http://dx.doi.org/10.1016/j.foodchem.2010.02.053

Martin, J.-C., et al. (2015). Can we trust untargeted metabolomics? Results of the metabo-ring initiative, a large-scale, multi-instrument inter-laboratory study. Metabolomics 11, 807-821 doi:10.1007/s11306-014-0740-0

Ragone, R., et al. (2015). Classification and chemometric study of Southern Italy monovarietal wines based on NMR and HPLC-DAD-MS. Food Science and Biotechnology 24, 817-826 doi:10.1007/s10068-015-0106-z

Salvatore, E., M. Cocchi, A. Marchetti, F. Marini, A. de Juan (2013). Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC-DAD and chemometric techniques. Analytica Chimica Acta 761, 34-45 doi:http://dx.doi.org/10.1016/j.aca.2012.11.015

Sanz-Cortes, M., R. J. Carbajo, F. Crispi, F. Figueras, A. Pineda-Lucena, E. Gratacos (2013). Metabolomic profile of umbilical cord blood plasma from early and late intrauterine growth restricted (IUGR) neonates with and without signs of brain vasodilation. PLoS One 8, e80121 doi:10.1371/journal.pone.0080121

Serrano-Lourido, D., J. Saurina, S. Hernández-Cassou, A. Checa (2012). Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. Food Chemistry 135, 1425-1431 doi:http://dx.doi.org/10.1016/j.foodchem.2012.06.010

Springer, A. E., J. Riedl, S. Esslinger, T. Roth, M. A. Glomb, C. Fauhl-Hassek (2014). Validated Modeling for German White Wine Varietal Authentication Based on Headspace Solid-Phase Microextraction Online Coupled with Gas Chromatography Mass Spectrometry Fingerprinting. Journal of Agricultural and Food Chemistry 62, 6844-6851 doi:10.1021/jf502042c

Versari, A., V. F. Laurie, A. Ricci, L. Laghi, G. P. Parpinello (2014). Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. Food Research International 60, 2-18 doi:http://dx.doi.org/10.1016/j.foodres.2014.02.007

Veselkov, K. A., et al. (2011). Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery. Analytical Chemistry 83, 5864-5872 doi:10.1021/ac201065j

Westerhuis, J., et al. (2008). Assessment of PLSDA cross validation. Metabolomics 4, 81-89 doi:10.1007/s11306-007-0099-6

Wolfender, J. L., S. Rudaz, Y. H. Choi, H. K. Kim (2013). Plant Metabolomics: From Holistic Data to Relevant Biomarkers. Current Medicinal Chemistry 20, 1056-1090

SUPPLEMENTARY INFORMATION

Table S1. List of reference standards and their retention times in both platforms used to build the RT correlation function shown in Fig 3.

| Compound	Platform 1	Platform 2	M.F.
Gallic acid	0.81	1.81	C7H6O5
3,4-dihidroxibenzoic acid	1.51	2.71	C7H6O4
Caftaric acid	1.88	2.4	C13H12O9
2,5-dihidroxibenzoic acid	2.43	3.01	C7H6O4
Chlorogenic acid	2.87	3.54	C16H18O9
Catechin	2.87	3.36	C15H14O6
Caffeic acid	3.06	4.06	C9H8O4
Procyanidin B2	3.19	3.53	C30H26O12
Epicatechin	3.32	4.16	C15H14O6
Syringic acid	3.15	4.27	C9H10O5
Ferulic acid	3.96	5.45	C10H10O4
trans-piceid acid	3.99	5.13	C20H22O8
Sinapic acid	3.99	5.47	C11H12O5
Quercetin-glucoside	4.04	6.67	C21H20O12
Quercitrin	4.38	7.69	C21H20O11
Myricetin	4.61	7.43	C15H10O8
trans-resveratrol	5.08	6.91	C14H12O3
Morin	5.11	8.25	C15H10O7
Quercetin	5.89	9.27	C15H10O7
Apigenin	6.99	11.14	C15H10O5
Kaempferol	7.11	10.88	C15H10O6
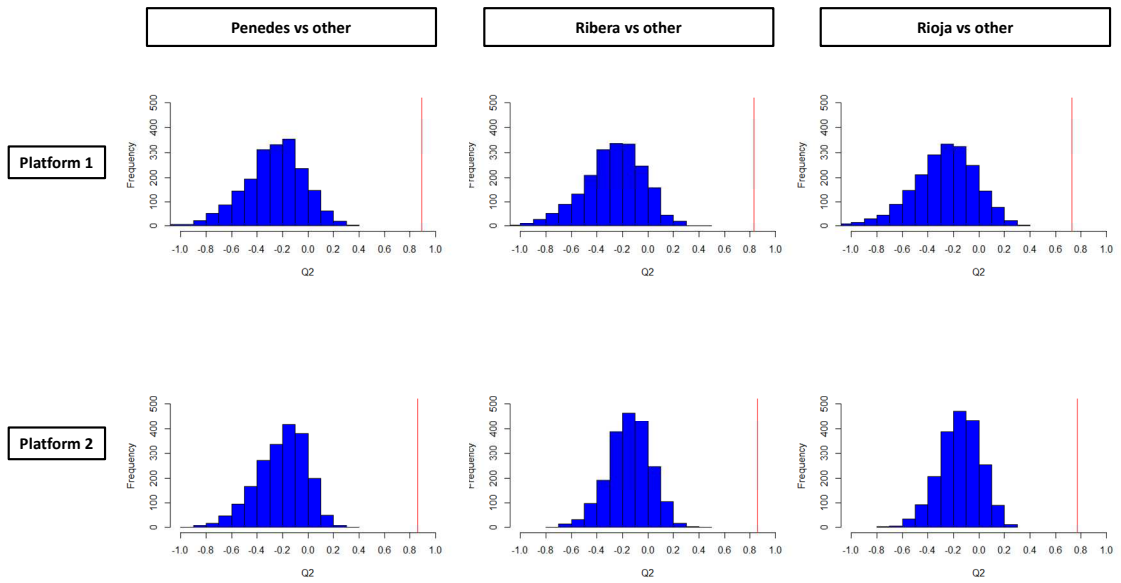
Fig. S1 Distribution of the Q2 values obtained after 2000 permutations of the original PDO classes for the different presented cases. Q2 values obtained for the non-permuted data

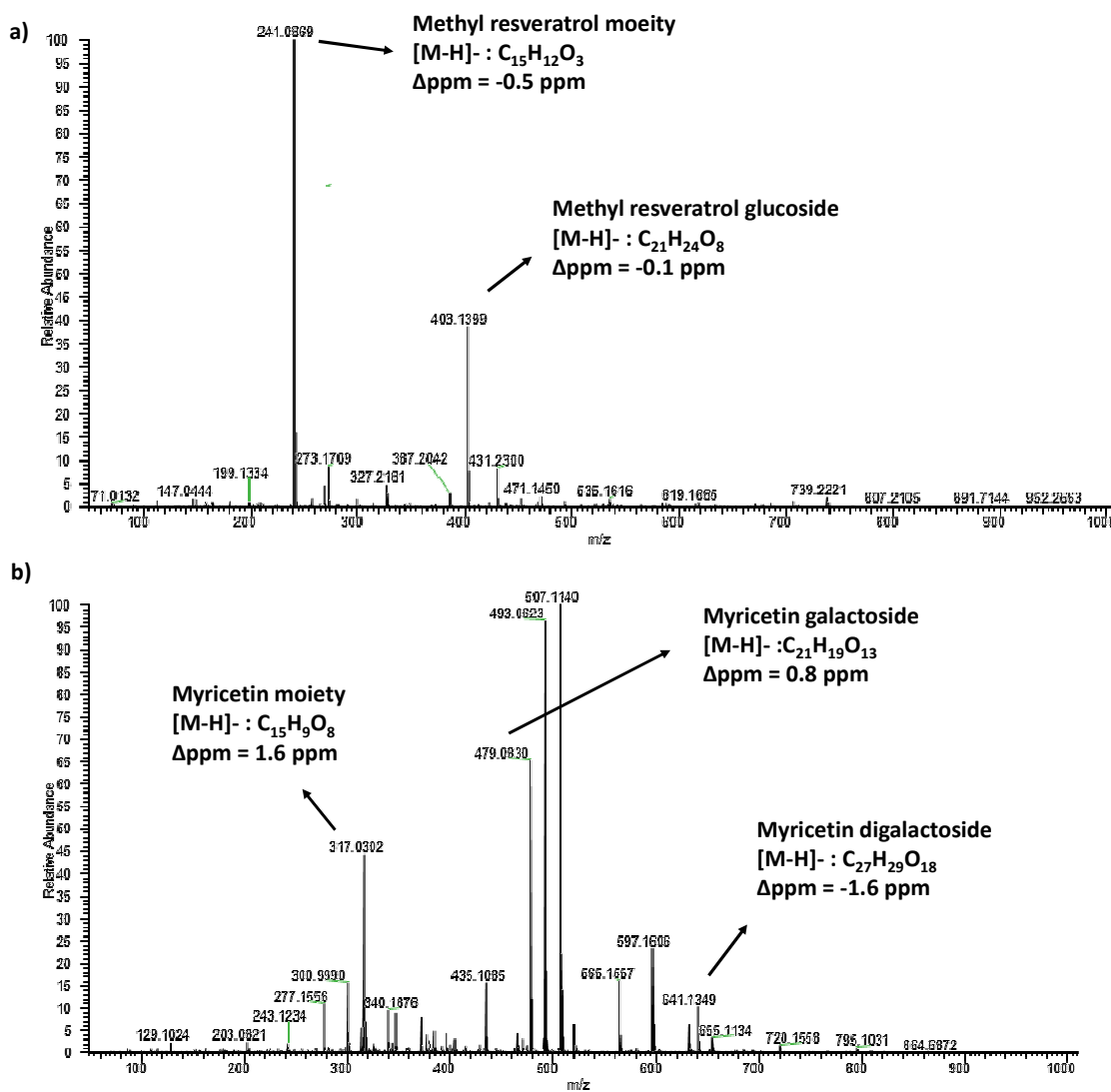(Table   1)   are   presented   as   vertical   red   lines   on   each   graph.

a)

Methyl resveratrol moeity
[M-H]- : $C_{15}H_{12}O_3$
Δppm = -0.5 ppm

Methyl resveratrol glucoside
[M-H]- : $C_{21}H_{24}O_8$
Δppm = -0.1 ppm

b)

Myricetin moiety
[M-H]- : $C_{15}H_9O_8$
Δppm = 1.6 ppm

Myricetin galactoside
[M-H]- :$C_{21}H_{19}O_{13}$
Δppm = 0.8 ppm

Myricetin digalactoside
[M-H]- : $C_{27}H_{29}O_{18}$
Δppm = -1.6 ppm

Fig. S2 Extracted MS spectrum (with background subtraction) for a) Penedes and b) Ribera del Duero wines from the chromatographic peaks observed at 6.02 and 3.72 minutes, respectively.

4.3.1. Discusión de resultados

A diferencia del artículo anterior, y como se puede observar en la figura 4.E. para el caso que nos ocupa, si se observó una elevada variabilidad en la mediana de las intensidades de los metabolitos encontrados, principalmente debida a una deriva instrumental al realizarse secuencias de inyección largas. Por este motivo, sí se efectuó una normalización Loess de las muestras previa al análisis estadístico de las mismas. Como se puede observar en la parte derecha de la figura 4.E., la normalización utilizada resulta muy eficaz, reduciendo la variación instrumental. Como se comentó anteriormente, esta aproximación podría conllevar la perdida de la varianza intrínseca de los metabolitos. No obstante, y a tenor de los resultados basándonos en el *fold change* de los marcadores encontrados, dicha normalización no alteró los datos de los marcadores.
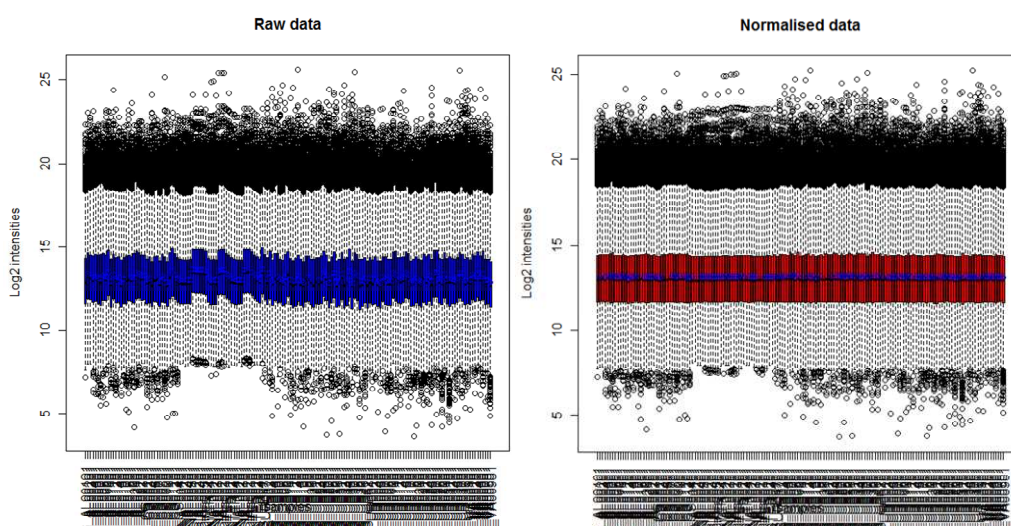

Figura 4.E. Box-plot de las variables (después de aplicar una trasformación logarítmica en base 2) de todas las muestras de vino inyectadas antes y después del proceso de normalización mediante el método Loess.

Por otra parte, aunque no se mencione en el artículo, la identificación tentativa de los marcadores encontrados se basó en la masa exacta y la información espectral obtenida mediante MS^E (i.e. sin necesidad de re-inyección). Para aquellos compuestos de los que se disponía de patrón de referencia, tanto los espectros de baja y alta energía como el tiempo de retención fueron usados para confirmar la identidad de los marcadores.

En el caso de la metabolómica en alimentos puede ocurrir la identificación de marcadores exógenos producto de las prácticas agrarias o de procesamiento de los alimentos. Este hecho puede resultar problemático y muy común y fue el caso en este artículo. Aunque no se menciona por no ser el objetivo de la investigación, en la segunda plataforma se identificó elel pesticida Metalaxyl. Uno de los marcadores encontrados en los datos obtenidos por ESI positivo fue el pico M280T671 cuya masa exacta era 280.1545. Este metabolito presentaba una mayor abundancia en la denominación Rioja (figura 4.F.) con un *fold change* de 13. Tras evaluar las posibles composiciones elementales para dicha masa, la formula molecular asignada fue C15H21NO4 (para la molécula neutra), con un error de masa de tan solo 0.2 mDa. Entre otras posibilidades, dicha fórmula molecular correspondía al fungicida Metalaxyl y los fragmentos encontrados se ajustaban tanto en masa como en abundancia a dicho pesticida. Posteriormente, y gracias a datos de patrones previamente inyectados, se pudo confirmar la identidad del fungicida en cuestión (figura 4.G.). Aunque el tiempo de retención no pudo compararse directamente, si se comparó el porcentaje necesario de fase móvil orgánica (en ambos casos MeOH) para la elución del analito, obteniendo una desviación únicamente del 2%. Aunque se observan algunas diferencias en cuanto a la abundancia relativa de los fragmentos, esto es debido a la formación del aducto sodiado en la muestra, el cual reduce el grado de fragmentación de la molécula.

Al igual que en el artículo anterior, el número limitado de muestras analizadas sugiere tratar a los compuestos seleccionados como candidatos. Sin embargo, este estudio pretendía profundizar en las diferencias de las dos plataformas aplicadas más que en la clasificación en sí. Sin embargo, aunque el total de muestras es pequeño, la

clasificación y al selección aleatoria de las muestras sugiere que un amplio estudio podría corroborar los resultados de éste.
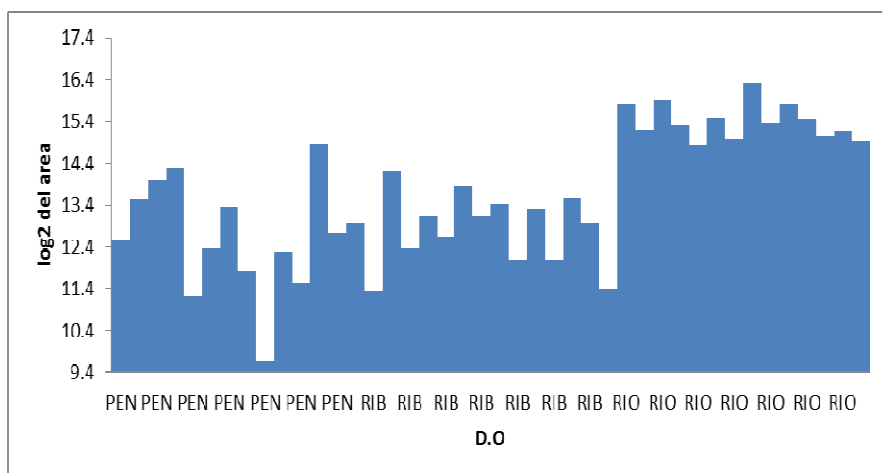

Figura 4.F. Logaritmo en base 2 del área del pico correspondiente al pesticida Metalaxyl en las diferentes muestras de vino analizadas según su D.O. PEN = Penedés, RIO = Rioja, Rib = Ribera del Duero.
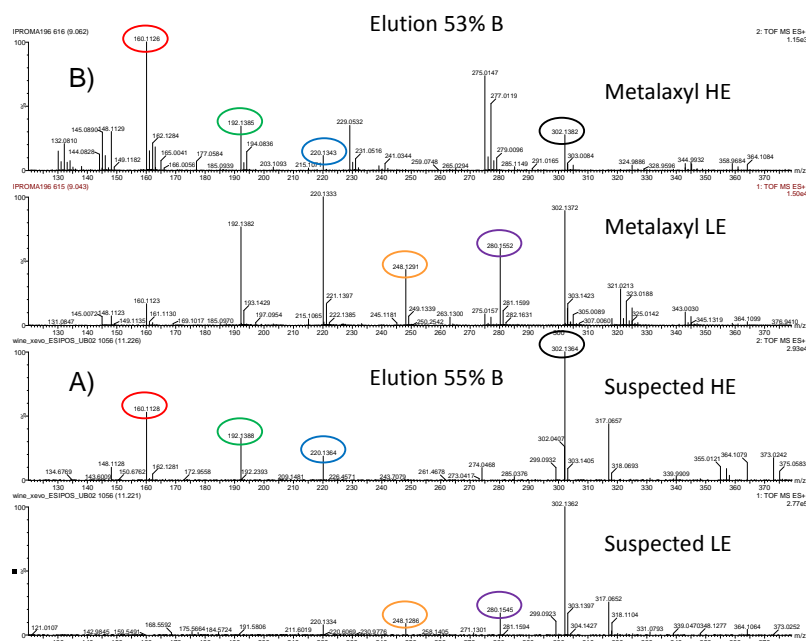

Figura 4.G. Espectros de baja y alta energía (LE y HE, respectivamente) del marcador elucidado como Metalaxyl (A) y el correspondiente patrón de referencia (B). Los iones no marcados en el espectro de HE en el patrón de referencia corresponden a un compuesto co-eluyente al Metalaxyl.

4.4. REFERENCIAS

Cotter, D., A. Maer, C. Guda, B. Saunders and S. Subramaniam (2006). "LMPD: LIPID MAPS proteome database." <u>Nucleic Acids Res</u> **34**(Database issue): D507-510.

Dunn, W. B., N. J. Bailey and H. E. Johnson (2005). "Measuring the metabolome: current analytical technologies." <u>Analyst</u> **130**(5): 606-625.

German, J. B., B. D. Hammock and S. M. Watkins (2005). "Metabolomics: building on a century of biochemistry to guide human health." <u>Metabolomics</u> **1**(1): 3-9.

Karpievitch, Y. V., A. R. Dabney and R. D. Smith (2012). "Normalization and missing value imputation for label-free LC-MS analysis." <u>BMC bioinformatics</u> **13 Suppl 16**.

Kopka, J., N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie and D. Steinhauser (2005). "GMD@CSB.DB: the Golm Metabolome Database." <u>Bioinformatics</u> **21**(8): 1635-1638.

Moco, S., L.-H. Tseng, M. Spraul, Z. Chen and J. Vervoort (2006). "Building-Up a Comprehensive Database of Flavonoids Based on Nuclear Magnetic Resonance Data." <u>Chromatographia</u> **64**(9-10): 503-508.

Peters, S., E. van Velzen and H. G. Janssen (2009). "Parameter selection for peak alignment in chromatographic sample profiling: objective quality indicators and use of control samples." <u>Anal Bioanal Chem</u> **394**(5): 1273-1281.

Smith, C. A., E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak (2006). "XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification." <u>Analytical Chemistry</u> **78**(3): 779-787.

Veselkov, K. A., L. K. Vingara, P. Masson, S. L. Robinette, E. Want, J. V. Li, R. H. Barton, C. Boursier-Neyret, B. Walther, T. M. Ebbels, I. Pelczer, E. Holmes, J. C. Lindon and J. K. Nicholson (2011). "Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery." <u>Analytical Chemistry</u> **83**(15): 5864-5872.

Want, E. (2009). "Challenges in applying chemometrics to LC-MS-based global metabolite profile data." <u>Bioanalysis</u> **1**(4): 805-819.

Want, E. J. and T. O. Metz (2010). MS Based Metabonomics. <u>Encyclopedia of Spectroscopy and Spectrometry</u>, Elsevier Ltd**:** 1663-1674.

Want, E. M., P. (2011). Processing and Analysis of GC/LC-MS-Based Metabolomics Data. <u>Metabolic Profiling, Methods in Molecular Biology</u>. S. S. B. Media.

Wishart, D. S. (2008). "Metabolomics: applications to food science and nutrition research." <u>Trends in Food Science and Technology</u> **19**(9): 482-493.

Wishart, D. S. (2008). "Metabolomics: applications to food science and nutrition research." <u>Trends in Food Science & Technology</u> **19**(9): 482-493.

Wishart, D. S., D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. Macinnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel and L. Querengesser (2007). "HMDB: the Human Metabolome Database." <u>Nucleic Acids Res</u> **35**(Database issue): D521-526.

Worley, B. and R. Powers (2013). "Multivariate Analysis in Metabolomics." <u>Current Metabolomics</u> **1**(1): 92-107.

CAPÍTULO 5

Conclusiones

La investigación realizada en esta tesis ha permitido el desarrollo y evaluación de una metodología analítica para el screening de un elevado numero de contaminantes organicos en muestras de diferente naturaleza. Dicha metodología permite el análisis cualitativo de contaminantes de un modo rápido y sencillo mediante la creación de bases de datos de compuestos, obteniendo así una visión más amplia sobre los posibles contaminantes presentes en la muestra. Por otro lado, la sensibilidad y amplia información suministrada por equipos de HRMS han permitido, en un estudio preliminar, mostrar diferencias en el perfil metabólico de pacientes con enfermedad pulmonar, mediante el análisis de muestras no invasivas pero de bajos niveles de concentración. Finalmente, se han encontrado potenciales marcadores capaces de clasificar alimentos en función de su origen gracias al análisis multivariante de los datos obtenidos mediante el acoplamiento LC-HRMS.

Además de esta conclusión general, se pueden establecer conclusiones específicas relacionadas con cada uno de los trabajos realizados:

1. Junto a la creación de librerías teóricas de compuestos de interés (básicamente consistentes en la fórmula empírica y la masa exacta), los nuevos avances en la instrumentación analítica en equipos de HRMS permiten crear amplias librerías empíricas de espectros, haciendo posible la detección e identificación de los compuestos en muy diversos tipos de muestras. Las diferentes funcionalidades instrumentales aseguran errores de masa por debajo de los 2 mDa, facilitando la identificación de los contaminantes de interés en un amplio rango de concentración.

2. La aproximación MS^E, y más concretamente aplicada mediante una rampa de energía de colisión, permite la obtención de espectros de masas muy similares

a los obtenidos mediante MS/MS, mostrando una gran robustez y suficiente grado de fragmentación para la mayoría de compuestos incluidos en la librería de espectros experimental. Además, mediante la adquisición en modo MS^E se facilita la confirmación de la identidad de los analitos y la elucidación de posibles desconocidos sin necesidad de re-inyección de la muestra.

3. La separación cromatográfica (usando condiciones genéricas y una columna de mayor longitud) y la calibración del eje de masas son factores claves en el análisis tipo screening pues permiten obtener mejor exactitud de masa en matrices complejas sin afectar a la sensibilidad o universalidad de la metodología.

4. La aproximación *post-target*, en la cual se investiga la presencia de un listado de compuestos una vez se han adquirido los datos en HRMS, comparada con el modo de trabajo *non-target* (no existe ningún tipo de selección de compuestos), ofrece resultados más fiables en términos de detección, además de ser más rápida y fácilmente automatizable. Para poder facilitar la confirmación de la identidad de forma simultánea a la detección, la base de datos teórica usada en este trabajo ha sido mejorada incluyendo no solo el tiempo de retención, sino la formula empírica de los principales fragmentos observados para aquellos compuestos de los que se disponía de patrón de referencia.

5. La aplicación de la metodología *post-target* ha permitido la identificación directa de un elevado número de compuestos de diferentes familias en muestras de origen ambiental, biológico y alimentario en aquellos casos en los que se disponía de patrón. Además, la gran cantidad de información obtenida ha facilitado la identificación tentativa de algunos contaminantes sin el uso de patrones de referencia. La confirmación de su identidad es posible en una etapa posterior mediante la inyección de los respectivos patrones, los cuales pueden ser adquiridos cuando ya existen evidencias sólidas sobre la presencia de los compuestos en las muestras.

6. La aplicación de métodos de screening de contaminantes orgánicos en aguas requiere de una pre-concentración de los analitos debido a la baja concentración habitualmente encontrada en las muestras. Con este objetivo, la técnica de SPE ha mostrado ser de fácil aplicación, con bajo consumo de disolventes y muy eficiente. La elección de un cartucho de extracción de base polimérica no-selectivo, como el Oasis HLB, ha sido una opción sencilla y fiable para el screening, y ha permitido extraer de forma efectiva los 120 compuestos usados en el test previo a la validación cualitativa, exceptuando el antibiótico Gabapectina.

7. Resulta indispensable realizar la validación cualitativa de la metodología de screening para el establecimiento de los límites de detección y de identificación. En nuestro caso, de los cerca de 150 compuestos modelo incluidos en la validación, el 88% fueron detectados y el 74% correctamente identificados al nivel de concentración más alto ensayado en aguas (1 µg/L). Al nivel más bajo (0.1 µg/L), el porcentaje de detección fue del 68%, el cual se redujo considerablemente, hasta un 38%, cuando se aplicaron las estrictas condiciones de identificación.

8. En estudios de amplio screening, como los realizados en esta Tesis, se ha puesto de manifiesto la necesidad de revisar los criterios usados en cuanto a la tolerancia máxima de las relaciones iónicas de intensidades (Q/q ratio) en comparación con los valores de referencia obtenidos con patrones. El no cumplimiento de las tolerancias (generalmente, entre 20 y 50% en función del valor de Q/q ratio) para las relaciones iónicas puede llevar a falsos negativos, por lo que sería conveniente revisar dichas tolerancias, o incluso establecer otros criterios basados en errores de masas, (junto al tiempo de retención) como condición para la identificación en el caso de aplicaciones no-dirigidas.

9. La metodología screening desarrollada no ha resultado eficiente para la detección de hormonas, debido a las condiciones de extracción y/o cromatográficas aplicadas, así como a la falta de sensibilidad del QTOF, insuficiente para la detección de estos compuestos en aguas. Por el contrario, para un importante número de analitos se pudieron detectar hasta 5

iones/fragmentos, medidos en masa exacta, incluso a niveles por debajo de los ensayados, lo cual aumenta la fiabilidad de las identificaciones.

10. El acoplamiento LC-HRMS se ha mostrado como una interesante alternativa a NMR para el análisis metabolómico de muestras en las que las concentraciones de los potenciales biomarcadores son bajas, como ha ocurrido en los trabajos realizados con los condensados de exhalados pulmonares.

11. El trabajo preliminar realizado ha demostrado el potencial de la técnica LC-HRMS para diferenciar la composición metabólica de condensados de exhalados pulmonares de pacientes control y con Enfermedad Pulmonar Obstructiva Crónica (EPOC), lo cual supone una gran ventaja al tratarse de una técnica no invasiva. Además, se propone el uso de condensadores sin partes reusables para evitar en la medida de lo posible la interferencia de contaminantes exógenos.

12. Las muestras de naranjas y vino de diferentes orígenes geográficos han podido ser separadas significativamente mediante análisis multivariante de los datos obtenidos por LC-HRMS y posteriormente procesados con XCMS. Dicha clasificación se corrobora gracias a la validación cruzada de los modelos estadísticos.

13. La normalización y control de calidad de los datos obtenidos mediante plataformas metabolómicas resultan de gran importancia dado el elevado número de muestras analizadas. Además del procesamiento posterior de datos, es necesario evaluar y simplificar lo máximo posible el tratamiento de la muestra con el fin de minimizar las posibles pérdidas de compuestos de interés.

14. La Citrusina D se muestra como un buen potencial marcador para la diferenciación de muestras de naranja de procedencia Valencia en comparación con aquellas del hemisferio sur (concretamente Argentina, Sud-África y Brasil). Se refleja además la importancia de validar los modelos y marcadores encontrados dado que cuando se tuvieron en cuenta muestras de cosechas consecutivas, algunos de los principales marcadores encontrados al analizar la primera cosecha fueron finalmente descartados.

15. Tras la aplicación de dos plataformas distintas para el análisis metabolómico de muestras de vino se concluye la importancia de la instrumentación y procesamiento de datos empleados para la obtención de marcadores. La utilización de diferentes condiciones cromatográficas, analizadores de masa (Orbitrap y QTOF) y parámetros de procesamiento de datos (entre otras variables) implica una distinta selección de marcadores. Sin embargo, ambas plataformas permiten la clasificación inequívoca de los vinos según su origen y producen resultados similares y/o complementarios en cuanto a los marcadores encontrados y las rutas metabólicas involucradas.

16. El análisis metabolómico puede verse influenciado por compuestos exógenos a la muestra, como ha ocurrido en esta Tesis en el caso del fungicida metalaxyl en las muestras de vino. Sin embargo, al mismo tiempo, el hecho de encontrar un plaguicida como marcador significativo y diferenciador entre vinos de distintas zonas geográficas aporta información añadida respecto a las prácticas agrarias en las diferentes zonas de cultivo.

CHAPTER 5
Conclusions

The research performed in this Thesis has allowed the development and evaluation of an analytical methodology for the screening of a high number of organic contaminants in different ature samples. This methodlogy permitted the qualitative analysis of pollutants in a fast and simple way thanks to the creation of a compound database, obtaining a wide vision of the contaminantion in the samples. On the other hand, the sensitivity and wide information supplied by HRMS allowed, I a preliminar study, to find differences in the metabolic profile of patients with pulmonary disease by the analysis of extremely low concentrated non-invase samples such as the exhaled breath condensates. Finally, potential markers for oranges and wines classification based on their geographical origin have been found by multivariate analysis from LC-HRMS data. In addition to this general conclusion, more specific conclusions can be established related to each work performed:

1. Together with the theoretical library of target compounds (consisting basically on a database with the compound name, empirical formula and exact mass), the recent advances in the analytical instrumentation allow the creation of highly robust empirical mass spectra libraries making feasible compound detection and identification in different sample types. The instrument functionalities ensure mass errors well below 2 mDa, facilitating the identification of the contaminants of interest in a wide range of concentrations.

2. The MS^E approach, and more concretely applied with a collision energy ramp, favours the acquisition of spectra similar to those obtained by MS/MS, showing a high robustness and sufficient fragmentation degree for most of the compounds included in the experimental spectra library. Furthermore, using MS^E acquisition mode, the confirmation of the identity of the analytes and the elucidation of

possible suspected or unknown compounds is facilitated without the need of reinjecting the sample.

3. The chromatographic separation (using generic conditions and increasing the column length) and the mass axis calibration are key factors in the wide-scope screening as they permit a mass accuracy improvement for complex matrices while not affecting the sensitivity or the scope range of the methodology.

4. The post-target approach, where the presence of a list of compounds is investigated after HRMS data acquisition, compared to the non-target one (no compound selection is performed a priori), offers more effective and reliable results in terms of detection and it results faster and easier to automate. In order to facilitate the detection and identification of the compounds, the data base has been improved including not only the retention time of the compounds but also the empirical formula of the main fragments for those analytes for which reference standard was available.

5. The application of the post-target methodology to biological, food and environmental samples has allowed the identification of an important amount of contaminants from different families. Furthermore, the information obtained has facilitated the tentative identification of some contaminants that have been subsequentely confirmed by the injection of the corresponding reference standard, which was acquired once solid evidences on their presence was established.

6. The application of the screening methodology for organic micro-contaminants in water requires a pre-concentration step due to the low concentration levels of the compounds usually found in the samples. With this aim, SPE has proven to be a simple, efficient and low solvent-consumption approach. The choice of generic polymeric SPE cartridges has shown to be reliable for screening purposes and has permitted the extraction of the entire 120 model compounds used for method evaluation, except the antibiotic Gabapentin.

7. A qualitative validation of the screening method is required in order to establish de limits of detection and identification. In our case, from the 150 model

compounds included in the validation, 88% were detected and 74% correctly identified at the highest concentration level tested in waters (1 μg/L). At the lowest level (0.1 μg/L), the percentage of detected compounds decreased down to 68%, and it was substantially reduced, down to 38%, when the strict identification conditions were applied.

8. After application of the wide-scope screening methods developed in this Thesis, it was deduced the necessity to review the criteria used for ion ratios tolerances (Q/q ratio) when compared with the reference values obtained by the injection of standards. The non-compliance of these tolerances (between 20 and 50% depending on the Q/q ratio) might lead to false negatives. Alternative criteria based on mass accuracy (together with retention time) or higher tolerances would be required for identification for non-target unknown compounds, or in post-target screening.

9. The screening methodology developed in this Thesis was not appropriate for the detection of the tested hormones surely due to the extraction and/or chromatographic conditions applied, as well as the lack of sensitivity of the QTOF MS, insufficient for the detection of these compounds in water. On the contrary, for an important number of analytes up to 5 ions/fragments could be used for detection and/or identification, even below the lowest concentration level tested. This fact improves significantly the reliability of the identifications.

10. LC-HRMS has been found to be an interesting alternative to NMR for the metabolomics analysis for samples with low biomarkers concentration levels, such as the exhaled breath condensates, due to the relevant information provided and the better sensitivity

11. The preliminary study performed in exhaled breath shows the ability of LC-HRMS to differentiate the metabolic composition of control samples and those from patients suffering from Chronic Obstructive Pulmonary Disease (COPD). This is an interesting approach as it implies the use of non-invasive sampling. Furthermore, the use of non-reusable condenser parts avoids the possible interference of exogenous contaminants coming from the cleaning steps.

12. Orange and wine samples from different geographical origins have been significantly classified by multivariate analysis using the data obtained from LC-HRMS analysis and subsequently processed by XCMS. This classification has been corroborated by cross-validation of the statistical models.

13. The normalization and quality control of the data obtained from metabolomics platforms may result crucial due to the high amount of samples analysed. However, additionally to data processing, it is necessary to evaluate and simplify the sample treatment in order to minimize possible metabolites loss.

14. Citrusin D seems to be a valid marker for the differentiation of orange samples from Valencia and those from South Hemisphere origin (concretely Argentine, South-Africa and Brazil). This work illustrates the importance of model validation as some of the compounds statistically significant for the model classification were proved to be not relevant after including data from a consecutive season (second year).

15. From the application of two different platforms for the metabolomics analysis of wine samples, the importance of the instrumentation and data treatment is concluded. The use of different chromatographic conditions, mass analyzers (Orbitrap and QTOF) and data processing parameters implies a different selection of the markers. Nevertheless, both platforms permit to classify unequivocally the wine samples tested according to their origin and both produce similar and/or complementary results regarding the biomarkers found and the involved metabolic pathways.

16. Metabolomics analysis can be influenced by exogenous compounds present in the samples, as occurred in this Thesis with the fungicide metalaxyl in wine samples. However, the fact of finding significant differences in these pesticide concentrations for different origins can highlight diverse agriculture practices in the different growing areas.