



UNIVERSIDAD
DE ALMERÍA

APLICACIÓN DE LA ESPECTROMETRÍA DE MASAS Y LA
RESONANCIA MAGNÉTICA NUCLEAR A LA
CARACTERIZACIÓN METABOLÓMICA DE BEBIDAS
ALCOHÓLICAS Y AL ANÁLISIS DE TÓXICOS EN AGUAS

*APPLICATION OF MASS SPECTROMETRY AND NUCLEAR
MAGNETIC RESONANCE TO THE METABOLOMIC
CHARACTERIZATION OF ALCOHOLIC BEVERAGES AND TO THE
ANALYSIS OF CONTAMINANTS IN WATER*

Tesis Doctoral

JOSÉ RAÚL BELMONTE SÁNCHEZ

Almería, Junio, 2020

MEMORIA

***Aplicación de la espectrometría de masas y la
resonancia magnética nuclear a la caracterización
metabolómica de bebidas alcohólicas y al análisis de
tóxicos en aguas***

Memoria presentada por

José Raúl Belmonte Sánchez

para optar al grado de Doctor en Ciencias Químicas

Programa de Doctorado en Química Avanzada

Departamento de Química y Física

Fdo. José Raúl Belmonte Sánchez

Director de Tesis

Codirector de Tesis

Fdo: José Luis Martínez Vidal

Fdo: Javier Arrebola Liébanas

AGRADECIMIENTOS

Siempre he considerado el avance del tiempo como una sucesión de eventos donde la dirección de los pasos define la ventura de cada uno. La intrepidez de cada acción en este viaje que es la vida da lugar al destino que te aguarda, a veces alcanzado con perseverancia, buenaventura y, en ocasiones, imprevisible.

Entre las líneas de la presente Tesis Doctoral, podréis discernir un intenso viaje lleno de decisiones. Una aventura imposible de consumir sin el apoyo de quienes me han rodeado. En cada esfuerzo, revés y frustración siempre he contado con el soporte de todos ellos. Y no solo eso, sino que en cada satisfacción, he contado con un festejo y júbilo memorables.

Quiero manifestar mi más sentido agradecimiento al grupo de profesionales y amigos de Química Analítica de Contaminantes por todo lo recibido a lo largo de este trayecto. Vosotros, quienes con vuestra incesante dirección, tutoría, aprecio y apoyo habéis hecho de mi camino un insólito viaje lleno de logros y satisfacción.

Me gustaría además hacer notorio el apoyo de mis padres quienes considero los verdaderos merecedores de ésta Tesis Doctoral, al haberme dado todas las herramientas necesarias para poder lograrlo. De igual forma, a la suerte de mi vida por ser quién es y verme como me ve. Familia, amigos, compañeros y aquellos que ya no están y tanto añoro: muchas gracias, no se expresa con palabras todo lo que suponéis día a día. Imposible de enumerar el continuo cúmulo de personas que han formado parte de ésta etapa.

No es mi intención expresar un sentimiento de desenlace en éste manifiesto, sino de continuidad, y la posibilidad de seguir contando con todos y cada uno de vosotros. Espero alentar a cada lector, sea concerniente o no de la presente Tesis Doctoral, a su lectura y así poder seguir caminando juntos en ésta sucesión de eventos que nos aguarda.

OBJETIVOS

El principal objetivo de esta Tesis, titulada “Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabólica de bebidas alcohólicas y al análisis de tóxicos en aguas”, está centrado en el desarrollo y validación de métodos analíticos, análisis metabólico e identificación de marcadores químicos de calidad en bebidas alcohólicas y aguas. A tal fin se han usado técnicas cromatográficas (cromatografía de gases, GC; cromatografía de líquidos, LC) acopladas a espectrometría de masas (MS) y de resonancia magnética nuclear (NMR).

Para lograr este objetivo general la presente Tesis se ha dividido en dos grandes bloques. El primer bloque se focaliza en la aplicación de técnicas de GC-MS, LC-MS y NMR para el estudio metabólico de ron, a fin de establecer métodos de control de calidad en el ámbito del marco alimentario de bebidas alcohólicas. Los estudios realizados han sido:

- Desarrollo de una estrategia de clasificación y discriminación de ron mediante análisis *fingerprinting* de compuestos volátiles/semivolátiles a través de GC-MS de triple cuadrupolo (QqQ). **Aportación científica 1**
- Desarrollo de una estrategia de clasificación y discriminación de ron mediante análisis *fingerprinting* de compuestos no volátiles a través de LC-MS de tipo Orbitrap (Orbitrap-MS). **Aportación científica 2**
- Aplicación de NMR para la clasificación de ron mediante análisis de los perfiles metabólicos de componentes mayoritarios. Adicionalmente se aplican estrategias de fusión de datos

combinando la información obtenida mediante las tres técnicas aplicadas: GC-MS, LC-MS y NMR. **Aportación científica 3**

Por otro lado, el segundo bloque incluye estudios relacionados con la aplicación analítica de la técnica de NMR para la determinación de tóxicos en aguas:

- Aplicación de la NMR de boro-11 para la determinación de ácido bórico en biocidas comerciales. **Aportación científica 4**
- Determinación de ácido etidróico en aguas industriales de lavado de vegetales mediante la NMR de fósforo-31. **Aportación científica 5**

RESUMEN

La presente Tesis Doctoral, titulada “*Aplicación de espectrometría de la masas y la resonancia magnética nuclear a la caracterización metabólica de bebidas alcohólicas y al análisis de tóxicos en aguas*”, estudia el uso de herramientas de la metabolómica en la clasificación e identificación de marcadores químicos de calidad en bebidas alcohólicas mediante técnicas espectrométricas (GC-MS, LC-MS) y de NMR. A tal fin, se han llevado a cabo distintas etapas del proceso analítico como la preparación de muestra, adquisición *fingerprinting*, pre-procesado de datos, modelado multivariante, validación de modelos y fusión de datos multi-técnica. Los modelos desarrollados se han aplicado a muestras reales de ron dorado para su clasificación en función de los procesos de preparación a los que se ven sometidos, así como a la determinación de marcadores químicos de calidad de éstos.

Además, se ha aplicado la técnica de NMR cuantitativa para la determinación de diferentes analitos. En concreto se han usado métodos de adquisición basados en núcleos magnéticamente activos de boro y fósforo para la cuantificación de ácido bórico y ácido etidróico respectivamente, tanto mediante cuantificación interna (estándar interno, IS, del núcleo observado) como externa (rectas de calibrado mediante uso de patrones analíticos). Los métodos se han validado según las guías vigentes para parámetros como selectividad, rango de trabajo, precisión interdía e intradía, límite de cuantificación (LOQ), recuperación e incertidumbre (U).

Así pues, los estudios realizados en la presente Tesis se pueden dividir en dos secciones: la primera orientada a la aplicación de herramientas de

la metabolómica y la segunda a la aplicación cuantitativa de la NMR. La primera sección cuenta con los siguientes trabajos:

- Desarrollo de una estrategia para la clasificación de ron usando tratamientos metabolómicos a través de huellas dactilares adquiridas por microextracción en fase sólida (SPME) de espacio en cabeza (HS) acoplada a cromatografía de gases–espectrometría de masas (HS-SPME-GC-MS). El método desarrollado se ha aplicado a rones dorados comerciales para su clasificación en función de las etapas de elaboración de éstos y determinación de marcadores químicos en su fase volátil y/o semivolátil.
- Estudio metabolómico de ron dorado mediante la adquisición de huellas dactilares de sus componentes no volátiles a través de inyección directa con dilución (*dilute and shoot*) y cromatografía de líquidos acoplada a espectrometría de masas de alta resolución (LC-HRMS). El método desarrollado ha permitido la clasificación y determinación de marcadores químicos de rones dorados comerciales.
- Estudio metabolómico de rones dorados mediante la adquisición de huellas dactilares a través resonancia magnética nuclear de protón ($^1\text{H-NMR}$) y fusión de datos multi-técnica (GC-MS, LC-MS y NMR). El modelo desarrollado mediante $^1\text{H-NMR}$ se ha aplicado al estudio de los compuestos mayoritarios presentes en rones dorados comerciales. Además, se han comparado las capacidades discriminatorias de las diferentes técnicas y su fusión en los análisis multivariantes aplicados.

Los trabajos que se incluyen en la segunda sección se muestran a continuación:

- Desarrollo y aplicación de un método mediante NMR de boro-11 ($^{11}\text{B-NMR}$) para la determinación de ácido bórico en biocidas sólidos. Los resultados se han comparado con técnicas previamente ya establecidas (espectrometría de masas por plasma de acoplamiento inductivo, ICP-MS). El método desarrollado se ha validado conforme a las guías vigentes y ha sido aplicado a diferentes muestras reales de biocidas comerciales para aguas agrícolas y de piscinas.
- Desarrollo y aplicación de un método mediante NMR de fósforo-31 ($^{31}\text{P-NMR}$) para la determinación de ácido etidrónico en agua de lavado de vegetales. Se han comparado distintos métodos de cuantificación con calibración interna y externa. El método desarrollado, y posteriormente validado, se ha aplicado al estudio de muestras de aguas provenientes del lavado de frutas y hortalizas.

ABSTRACT

This Doctoral Thesis entitled “Application of mass spectrometry and nuclear magnetic resonance to the metabolomic characterization of alcoholic beverages and to the analysis of contaminants in water” studies the application of metabolomic tools in the classification and identification of quality chemical biomarkers in alcoholic beverages using spectrometric techniques (GC-MS, LC-MS) and NMR, responding to the current demand for control and monitoring in this field. To this end, different stages have been carried out such as sample preparation, fingerprinting acquisition, data pre-processing, multivariate analysis, model validation and multi-technical data fusion. The developed methodologies have been applied to real samples of golden rum for its classification based on the preparation processes, as well as the determination of chemical quality markers.

In addition, the quantitative NMR technique has been applied for the determination of different analytes. In particular, acquisition methods based on magnetically active boron and phosphorus nuclei have been used for the quantitation of the analytes by internal quantitation (internal standard, IS, of the observed nucleus) and external quantitation (calibration using analytical standards). The methods have been validated according to current guidelines for parameters such as selectivity, working range, interday and intraday accuracy, limit of quantitation (LOQ), recovery and uncertainty (U).

Thus, the studies carried out in this thesis can be divided into two sections: the first focused on the application of metabolomic tools and the second to the quantitative application of NMR. The first section has the following works:

- Development of a strategy for the classification of rum using metabolomic treatments through fingerprints acquired by head space (HS) and solid phase microextraction (SPME) coupled to gas chromatography – mass spectrometry (HS-SPME-GC- MS). The methodology developed has been applied to commercial golden rums for classification according to the stages of their preparation and determination of chemical biomarkers in their volatile and/or semi-volatile phase.
- Metabolomic study of golden rum by fingerprint acquisition of its non-volatile components through direct injection (dilute and shoot) and liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS). The methodology developed has allowed the classification and determination of chemical biomarkers of commercial golden rums.
- Metabolomic study of golden rums by acquiring fingerprints through proton nuclear magnetic resonance ($^1\text{H-NMR}$) and multi-technical data fusion (GC-MS, LC-MS and NMR). The methodology developed has been applied to the study of the majority compounds present in commercial golden rums. In addition, the discriminatory capacities of the different techniques and their fusion in the multivariate analyses applied have been compared.

The works included in the second section are shown below:

- Development and application of a boron-11 NMR detection method ($^{11}\text{B-NMR}$) for the determination of boric acid in solid biocides. The results have been compared with previously established techniques (ICP-MS). The developed method has

been validated in accordance with current guidelines and has been applied to different real samples of commercial tablets for agricultural waters and swimming pools.

- Development and application of a phosphorus-31 NMR detection method (^{31}P -NMR) for the determination of etidronic acid in vegetable washing water. Different quantification methods have been compared with internal and external calibration. The method developed, and subsequently validated, has been applied to the study of real samples of water from the washing of vegetables and fruits.

LISTADO DE PUBLICACIONES INCLUIDAS EN LA PRESENTE TESIS DOCTORAL

Aportación científica 1. Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

J.R. Belmonte-Sánchez, S. Gherghel, J. Arrebola-Liébanas, R.R. González, J.L.M. Vidal, I. Parkin, A.G. Frenich. *Talanta*. 187 (2018) 348–356.

Aportación científica 2. An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography–Orbitrap mass spectrometry and chemometric strategies

J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido Frenich. *J. Agric. Food Chem.* 67 (2019) 1302–1311.

Aportación científica 3. ^1H -NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis

J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido Frenich. *Food Chemistry* 137 (2020) 126363.

Aportación científica 4. Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides

L.M. Aguilera-Sáez, J.R. Belmonte-Sánchez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich, I. Fernández. *Analyst*. 143 (2018), 4707-4714.

Aportación científica 5. Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative ^{31}P nuclear magnetic resonance method

J.R. Belmonte-Sánchez, L.M. Aguilera-Sáez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich. *Microchem. J.* 150 (2019) 104083.

CONTRIBUCIÓN DEL AUTOR EN CADA PUBLICACIÓN

Aportación Científica 1 → José Raúl Belmonte Sánchez participó en la revisión bibliográfica, implementación del *software* usado, diseño de los experimentos, realización de la parte experimental y en la redacción del manuscrito.

Aportación Científica 2 → José Raúl Belmonte Sánchez participó en la revisión bibliográfica, implementación del *software* usado, diseño de los experimentos, realización de la parte experimental y en la redacción del manuscrito.

Aportación Científica 3 → José Raúl Belmonte Sánchez participó en la revisión bibliográfica, implementación del *software* usado, diseño de los experimentos, realización de la parte experimental y en la redacción del manuscrito.

Aportación Científica 4 → José Raúl Belmonte Sánchez participó en la revisión bibliográfica, diseño de los experimentos, realización de la parte experimental y en la redacción del manuscrito.

Aportación Científica 5 → José Raúl Belmonte Sánchez participó en la revisión bibliográfica, diseño de los experimentos, realización de la parte experimental y en la redacción del manuscrito.

ABREVIATURAS Y ACRÓNIMOS

APCI	Ionización química a presión atmosférica (<i>Atmospheric Pressure Chemical Ionization</i>)
AUC	Área bajo la curva (<i>Area Under the Curve</i>)
B ₀	Campo magnético externo
CC%	Valor de muestras correctamente clasificadas (<i>Correctly Classified samples</i>)
CI	Ionización química (<i>Chemical Ionization</i>)
CV	Validación cruzada (<i>Cross Validation</i>)
CV-ANOVA	Análisis de la varianza mediante validación cruzada (<i>Cross Validated ANalysis Of VAriance</i>)
D ₁	Tiempo de relajación
DNA	Ácido Desoxirribonucleico (<i>Deoxyribonucleic acid</i>)
EAA	Espectroscopía de Absorción Atómica
EI	Ionización electrónica (<i>Electron Ionization</i>)
ESI	Ionización por electronebulización (<i>ElectroSpray Ionization</i>)
FDA	Administración de Medicamentos y Alimentos (<i>Food and Drug Administration</i>)
FTIR	Espectroscopía de transmisión de infrarrojo con transformada de Fourier (<i>Fourier-Transform InfraRed spectroscopy</i>)
GC	Cromatografía de gases (<i>Gas Chromatography</i>)
HCA	Análisis de agrupamiento jerárquico (<i>Hierarchical Clustering Analysis</i>)
HPLC	Cromatografía de líquidos de alta eficacia (<i>High Performance Liquid Chromatography</i>)

HRF	Valor de filtrado de alta resolución (<i>High-Resolution Filtering score</i>)
HRMS	Espectrometría de masas de alta resolución (<i>High Resolution Mass Spectrometry</i>)
HS	Espacio en cabeza (<i>Head Space</i>)
I	Número cuántico de espín nuclear
IC	Cromatografía iónica (<i>Ion Chromatography</i>)
ICP-MS	Espectrometría de Masas con Plasma de Acoplamiento Inductivo (<i>Inductively Coupled Plasma Mass Spectrometry</i>)
IRM	Resonancia Magnética de Imágenes (<i>Magnetic Resonance Imaging</i>)
IS	Estándar interno (<i>Internal Standard</i>)
IUPAC	Unión internacional de química pura y aplicada (<i>International Union of Pure and Applied Chemistry</i>)
K	Número de variables de un modelo multivariante
KI	Índice de Retención de Kovats
LC	Cromatografía de líquidos (<i>Liquid Chromatography</i>)
LDA	Análisis discriminante lineal (<i>Linear Discriminant Analysis</i>)
LOD	Límite de detección (<i>Limit Of Detection</i>)
LOQ	Límite de cuantificación (<i>Limit Of Quantitation</i>)
LRMS	Espectrometría de masas de baja resolución (<i>Low Resolution Mass Spectrometry</i>)
M	Masa molecular

MIR	Infrarrojo medio (<i>Mid Infra-Red</i>)
MRM	Monitorización de reacción múltiple (<i>Multi-Reaction Monitoring</i>)
MS	Espectrometría de masas (<i>Mass Spectrometry</i>)
MS/MS	Espectrometría de masas en tándem (<i>Tandem mass spectrometry</i>)
N	Número de observaciones de un modelo multivariante
NIR	Infrarrojo cercano (<i>Near Infra-Red</i>)
NMR	Resonancia magnética nuclear (<i>Nuclear Magnetic Resonance</i>)
NOE	Efecto nuclear Overhauser (<i>Nuclear Overhauser Effect</i>)
NS	Número de <i>scans</i>
OPLS-DA	Análisis discriminante por mínimos cuadrados parciales ortogonales (<i>Orthogonal Partial least Squares-Discriminant Analysis</i>)
P	Valor de significancia de la variable (<i>p-value</i>)
P ₁	Pulso magnético aplicado
PC	Componente principal (<i>Principal Component</i>)
PCA	Análisis de componentes principales (<i>Principal Component Analysis</i>)
PLS-DA	Mínimos cuadrados parciales-análisis discriminante (<i>Partial Least Squares Discriminant Analysis</i>)
ppm	Partes por millón
PQM	Proceso Químico de Medida
Q	Momento eléctrico cuadrupolar (<i>Quadrupole</i>)

QC	Control de calidad (<i>Quality Control</i>)
q NMR	Resonancia Magnética Nuclear cuantitativa (<i>Quantitative Nuclear Magnetic Resonance</i>)
QqQ	Triple cuadrupolo (<i>Triple Quadrupole</i>)
Q-TOF	Sistema híbrido cuadrupolo/tiempo de vuelo (<i>Quadrupole-Time Of Flight</i>)
R	Coefficiente de correlación
RG	Ganancia del receptor (<i>Receiver Gain</i>)
RMSEC	Raíz del error cuadrático medio de calibración (<i>Root Mean Square Error of Calibration</i>)
RMSECV	Raíz del error cuadrático medio de la validación cruzada (<i>Root Mean Square Error of Cross-Validation</i>)
RMSEP	Raíz del error cuadrático medio de predicción (<i>Root Mean Square Error of Prediction</i>)
ROC	Característica de funcionamiento del receptor (<i>Receiver Operating Characteristic</i>)
RSD	Desviación estándar relativa (<i>Relative Standard Deviation</i>)
RT	Tiempo de retención (<i>Retention Time</i>)
S/N	Relación señal/ruido (<i>Signal-to-Noise ratio</i>)
SPME	Microextracción en fase sólida (<i>Solid Phase MicroExtraction</i>)
T_1	Tiempo de relajación longitudinal
TMS	Tetrametilsilano- d_4
TMSP	3-[Trimetilsilil]-propionato- d_4 de sodio
UV-Vis	Ultravioleta-visible (<i>UltraViolet-VISible</i>)

VIP	Importancia de la variable en la proyección (<i>Variable Importance in Projection</i>)
δ	Desplazamiento químico

ÍNDICE DE CONTENIDO

	Página
1 CAPÍTULO PRIMERO. MARCO TEÓRICO	7
1.1 Contextualización.....	7
1.2 Metabolómica	13
1.2.1 <i>Introducción al análisis metabolómico</i>	<i>13</i>
1.2.2 <i>Adquisición de datos</i>	<i>19</i>
1.2.3 <i>Pretratamiento de señales instrumentales</i>	<i>21</i>
1.2.4 <i>Métodos multivariantes.....</i>	<i>29</i>
1.2.5 <i>Validación de modelos multivariantes</i>	<i>34</i>
1.2.6 <i>Marcadores metabolómicos</i>	<i>38</i>
1.2.7 <i>Fusión de datos</i>	<i>43</i>
1.3 qNMR.....	46
1.3.1 <i>Introducción a la NMR cuantitativa.....</i>	<i>46</i>
1.3.2 <i>Proceso analítico qNMR: pretratamiento y adquisición.....</i>	<i>56</i>
1.3.3 <i>Validación del método.....</i>	<i>60</i>
1.4 Bibliografía.....	65
2 CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON	89
2.1 Introducción	89
2.2 Bibliografía.....	94
2.3 Aportación científica 1. Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry	97
2.4 Aportación científica 2. An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography-orbitrap mass spectrometry and chemometric strategies	107
2.5 Aportación científica 3. ¹H-NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis.....	119

3	CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA DETERMINACIÓN DE TÓXICOS.....	131
3.1	<i>Introducción</i>	<i>131</i>
3.2	<i>Bibliografía.....</i>	<i>135</i>
3.3	<i>Aportación científica 4. Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides.....</i>	<i>137</i>
3.4	<i>Aportación científica 5. Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative³¹P nuclear magnetic resonance method.....</i>	<i>147</i>
4	CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA.....	157
4.1	<i>Introducción</i>	<i>157</i>
4.2	<i>Aplicación de herramientas de la metabolómica al estudio de ron</i>	<i>158</i>
4.2.1	<i>Adquisición de rones</i>	<i>158</i>
4.2.2	<i>Análisis de la fracción volátil.....</i>	<i>160</i>
4.2.3	<i>Análisis de la fracción no volátil.....</i>	<i>165</i>
4.2.4	<i>Análisis mediante NMR.....</i>	<i>167</i>
4.2.5	<i>Estudio metabolómico de rones</i>	<i>168</i>
4.2.6	<i>Fusión y comparación de datos multi-técnica.....</i>	<i>187</i>
4.3	<i>Aplicación de la qNMR a la determinación de compuestos tóxicos</i>	<i>194</i>
4.3.1	<i>Determinación de ácido bórico en biocidas comerciales</i>	<i>194</i>
4.3.2	<i>Determinación de ácido etidróico en aguas de lavado de vegetales</i>	<i>199</i>
5	CAPÍTULO QUINTO. CONCLUSIONES	205
6	CAPÍTULO SEXTO. COPYRIGHT	207
7	CAPÍTULO SÉPTIMO. OTRAS APORTACIONES CIENTÍFICAS	217



ÍNDICE DE TABLAS

<i>Tabla 1. Términos y definiciones comunes en los estudios metabolómicos</i>	8
<i>Tabla 2. Términos y definiciones comunes en la quimiometría aplicada a la metabolómica</i>	17
<i>Tabla 3. Listado de software disponible para deconvolución en MS</i>	23
<i>Tabla 4. Términos y definiciones comunes en el pretratamiento de datos de estudios metabolómicos</i>	27
<i>Tabla 5. Herramientas para el análisis estadístico de datos en estudios metabolómicos</i>	29
<i>Tabla 6. Denominación de análisis y términos en modelos multivariantes</i>	32
<i>Tabla 7. Bases de datos más utilizadas en la identificación tentativa de los analitos</i>	40
<i>Tabla 8. Ventajas y desventajas de las principales técnicas de análisis cuantitativo</i>	53
<i>Tabla 9. Principales consideraciones a tener en cuenta en determinaciones mediante qNMR</i>	60
<i>Tabla 10. Revisión bibliográfica sobre estudios de ron</i>	92
<i>Tabla 11. Etiquetado de botellas e información del fabricante para las muestras de ron analizadas</i>	161
<i>Tabla 12. Parámetros y condiciones para el análisis metabolómico de rones mediante HS-SPME-GC-MS</i>	165
<i>Tabla 13. Parámetros y condiciones para el análisis metabolómico de rones mediante LC-Orbitrap</i>	166
<i>Tabla 14. Parámetros y condiciones para el análisis metabolómico de rones mediante NMR</i>	167
<i>Tabla 15. Compuestos tentativamente identificados como más discriminantes en el estudio metabolómico de rones mediante HS-SPME-GC-MS y datos obtenidos para su identificación</i>	174
<i>Tabla 16. Compuestos identificados tentativamente como más discriminantes en el conjunto de rones mediante LC-Orbitrap-MS</i>	184

<i>Tabla 17. Comparación de los resultados multivariantes obtenidos para los distintos tipos de adquisición mediante NMR.....</i>	<i>188</i>
<i>Tabla 18. Valores de validación obtenidos para el análisis metabólico de las distintas técnicas individualmente y mediante fusión de datos de nivel bajo y medio.....</i>	<i>193</i>
<i>Tabla 19. Parámetros y condiciones para la determinación de ácido bórico mediante NMR.....</i>	<i>196</i>
<i>Tabla 20. Parámetros de validación obtenidos para el método de ¹¹B-NMR propuesto.</i>	<i>199</i>
<i>Tabla 21. Parámetros y condiciones para la determinación de ácido etidróico mediante NMR.....</i>	<i>200</i>
<i>Tabla 22. Parámetros de validación obtenidos para el método de ³¹P-NMR propuesto.</i>	<i>202</i>



ÍNDICE DE FIGURAS

<i>Figura 1. Flujo de trabajo general en un estudio metabolómico</i>	<i>15</i>
<i>Figura 2. Esquema general de un equipo de resonancia magnética nuclear moderno</i>	<i>46</i>
<i>Figura 3. Esquema de los distintos parámetros de adquisición, tratamientos de pre-procesado y aplicaciones posibles que requiere un análisis de qNMR</i>	<i>47</i>
<i>Figura 4. Resultados de búsqueda en Scopus del número de publicaciones por año para las palabras clave “quantitative NMR”, "q-NMR" y "qNMR"</i>	<i>49</i>
<i>Figura 5. Resultados de búsqueda en Scopus del número de publicaciones por año para las palabras clave “Heteronuclear quantitative NMR”, " Heteronuclear q-NMR" y " Heteronuclear qNMR"</i>	<i>51</i>
<i>Figura 6. Diagrama de flujo de elaboración de ron</i>	<i>90</i>
<i>Figura 7. Estructura química del ácido bórico (H_3BO_3).</i>	<i>132</i>
<i>Figura 8. Estructura química del ácido etidróico ($C_2H_8O_7P_2$).</i>	<i>133</i>
<i>Figura 9. Ejemplo de etiquetado de un ron dorado comercial</i>	<i>159</i>
<i>Figura 10. Cromatograma obtenido del análisis de una muestra de ron dorado mediante HS-SPME-GC-MS.....</i>	<i>164</i>
<i>Figura 11. Cromatograma obtenido del análisis de ron dorado mediante LC-HRMS (Orbitrap-MS) en modo de ionización negativo (arriba) y positivo (abajo)</i>	<i>166</i>
<i>Figura 12. Espectro de NMR de un ron dorado entre 6.0-0.0 ppm: a) sin supresión ni desacoplamiento de señales, b) con supresión de señales y sin desacoplamiento de carbono y c) con supresión de señales y desacoplamiento de carbono</i>	<i>168</i>
<i>Figura 13. Dendograma obtenido mediante HCA del conjunto total de rones</i>	<i>170</i>
<i>Figura 14. Representación PCA del conjunto de rones basado en aquellas variables más discriminantes</i>	<i>171</i>
<i>Figura 15. Representación LDA del conjunto de rones basado en a) Origen, b) Materia prima, c) Destilación, d) Barril y e) Edad.....</i>	<i>173</i>

Figura 16. Representación de barras de la suma de etíl ésteres en rones e indicaciones de tendencias mediante flechas 177

Figura 17. Representación tridimensional del PCA obtenido para el conjunto de rones mediante LC-MS..... 179

Figura 18. Representación gráfica de Hotelling's T^2 (arriba) y $D_{mod}X$ (abajo) junto a los outliers detectados (límites de confianza= 95-99%) 179

Figura 19. Representación PLS-DA basada en LC-MS del conjunto de rones basado en: a) Origen, b) Materia prima, c) Barril y d) Edad 181

Figura 20. Representación del PCA obtenido del análisis del conjunto de rones mediante NMR 190

Figura 21. Representación PLS-DA basada en 1H -NMR del conjunto de rones basado en: a) Edad, b) Barril, c) Destilación y d) Materia prima 191

Figura 22. Representación PLS-DA basada en fusión de datos de nivel bajo del conjunto de rones basado en: a) Edad, b) Barril, c) Destilación y d) Materia prima..... 195

Figura 23. Influencia de: a) Número de scans en la señal absoluta, b) Número de scans en la relación S/N, c)Tiempo de adquisición y d)Tiempo de relajación en la señal absoluta..... 197

Figura 24. a)Linealidad obtenida para el ácido bórico en agua y b)Espectros de ^{11}B -NMR a lo largo del rango de trabajo 198

Figura 25. Determinación del tiempo de relajación óptimo mediante las señales de fósforo-31 del ácido etidróico y del IS 201

Figura 26. Influencia del número de scans sobre la integral absoluta del ácido etidróico 202



ÍNDICE DE ECUACIONES

<i>Ecuación 1. Método de escalado mediante centrado medio</i>	<i>24</i>
<i>Ecuación 2. Método de transformación de escalado unitario (autoescalado)</i>	<i>25</i>
<i>Ecuación 3. Método de transformación de escalado Pareto.....</i>	<i>25</i>
<i>Ecuación 4. Método de transformación de escalado por rango</i>	<i>26</i>
<i>Ecuación 5. Método de transformación logarítmica.....</i>	<i>26</i>
<i>Ecuación 6. Método de transformación de potencia.....</i>	<i>27</i>
<i>Ecuación 7. Ecuación para el cálculo de la raíz del error cuadrático medio</i>	<i>38</i>
<i>Ecuación 8. Ecuación para el cálculo de la concentración del analito objetivo en valores absolutos mediante qNMR y calibración interna.</i>	<i>58</i>

ÍNDICE DE ECUACIONES



1 CAPÍTULO PRIMERO. MARCO TEÓRICO

1.1 Contextualización

La disponibilidad de recursos analíticos bien establecidos hoy en día permite el desarrollo de nuevas formas de afrontar retos comunes del ámbito de la seguridad agroalimentaria. En concreto, uno de los principales problemas de los últimos años es el control de calidad tanto de alimentos como de productos, con impactos económicos, sociales, de salud y ambientales. Las ciencias ómicas, que permiten el estudio de un gran número de moléculas de un organismo, están avanzando en las últimas décadas como una de las mejores opciones para dar respuesta a dichos problemas.

Existen tres disciplinas que pueden englobarse dentro de las ciencias ómicas, cada una con referencia a diferentes técnicas en el campo de la biotecnología [1]. La primera de ellas es la genómica, encargada del estudio del ácido desoxirribonucleico (DNA) presente en los organismos. La proteómica se basa en el estudio del proteoma, es decir el conjunto de todas las proteínas presentes en las células, así como las posibles modificaciones que puedan sufrir en su ciclo vital. Por último, la metabolómica trata del estudio de todos los metabolitos celulares, es decir, de todas aquellas moléculas implicadas en las rutas de transformación de diversos productos celulares. El conjunto de todos los metabolitos conforma el denominado metaboloma. Recientemente, el desarrollo y especialización en el campo de las disciplinas ómicas ha dado lugar a la aparición de nuevas subdisciplinas como la epigenómica,

lipidómica o la transcriptómica [2,3]. Los términos más comunes usados en los estudios metabolómicos y sus definiciones se encuentran recogidos en Tabla 1.

Tabla 1. Términos y definiciones comunes en los estudios metabolómicos

Término	Definición	Referencia
Biomarcador	Compuesto químico analizado en un material biológico relativo a una exposición, susceptibilidad o efecto biológico	[4]
Huella dactilar metabolómica	Perfil obtenido de un análisis que proporciona información de los analitos presentes en la matriz de estudio (“ <i>metabolic fingerprint</i> ”)	[5,6]
Metaboloma	Conjunto de moléculas de bajo peso molecular (<1500 Daltons) que se encuentran en el interior, o son secretadas por una célula o tejido de una matriz biológica	[5,6]
Metabolómica	Análisis completo del metaboloma de un sistema biológico en unas condiciones concretas. No tiene por qué ser cuantitativo o identificativo de los analitos	[5,6]
Metabolómica dirigida (target)	Análisis del metaboloma centrado en una lista específica de analitos, generalmente con un interés común	[7]
Metabolómica no dirigida (untarget)	Análisis global del metaboloma con el objetivo de estudiar tantos analitos como sea posible	[7]
Ómica	Estudio de conjuntos relacionados de moléculas orgánicas de manera íntegra (p. ej. genómica, transcriptómica, proteómica, etc.)	[8]
Perfil metabolómico	Análisis y estudio de un número limitado y predeterminado de analitos relacionados entre sí (por presentar similitudes físico-químicas o por participar en una ruta metabólica concreta)	[9]



En los últimos años, la metabolómica se ha reconocido como una de las disciplinas más prometedoras en el campo de las ciencias ómicas, en diferentes áreas como el control de calidad y la identificación de marcadores en matrices alimentarias [10–12]. Algunas de las estrategias más ampliamente usadas para dar solución a los problemas de seguridad alimentaria son los análisis de perfiles metabolómicos y la determinación de huellas dactilares [13]. Una de las principales diferencias entre el análisis convencional y el análisis metabolómico es el número de analitos cuyo estudio se aborda.

Las investigaciones enfocadas al estudio y comprensión del metaboloma en áreas como enfermedades humanas, análisis de plantas o alimentos son claros ejemplos de impacto y crecimiento de las ciencias ómicas. En concreto, el estudio de alimentos se encuentra en auge debido a la cada día mayor demanda de los consumidores de productos de alta calidad. Este hecho se ve reflejado en la abundancia de estudios metabolómicos que se han realizado en la última década en numerosos alimentos y bebidas como café [13,14], té [15], aceite [16,17], queso [18], azúcar de caña [19], miel [20], aceite de palma [21], etc. También existen un diferentes estudios sobre bebidas alcohólicas como cerveza [22,23] o vino [24–26], entre otras.

La calidad final de este tipo de productos puede depender de diversas variables: origen, composición, propiedades físicas, evaluación sensorial satisfactoria, procesamiento y almacenamiento del producto, etc. Cabe indicar que, en cuanto a la calidad de los alimentos, ésta puede evaluarse no sólo desde la perspectiva de la presencia/concentración de componentes beneficiosos para la salud y que incrementan la calidad del

CAPÍTULO PRIMERO. MARCO TEÓRICO

producto, sino también desde la perspectiva de la identificación de compuestos cuya presencia disminuyan su calidad (sustitución por ingredientes más económicos, diluciones, descripción incompleta del producto en su etiquetado, etc.).

Los últimos avances en las técnicas analíticas han sido cruciales para el desarrollo de métodos que puedan abarcar todos los requisitos actuales en materia de la calidad alimentaria. Hoy en día se dispone de diferentes técnicas con capacidad de obtener gran cantidad de información de las matrices alimentarias. Entre ellas destacan la cromatografía de gases acoplada a espectrometría de masas (GC-MS), cromatografía de líquidos acoplada a espectrometría de masas (LC-MS) o la resonancia magnética nuclear (NMR) [27–29]. Una ventaja añadida de éste tipo de técnicas es su versatilidad para ser usadas con fines cuantitativos.

La elección de la técnica más adecuada dependerá fundamentalmente de las características físico-químicas de los componentes a determinar en la muestra, como pueden ser su polaridad, volatilidad o nivel de concentración. En el caso de compuestos que presentan una polaridad baja y volatilidad alta-media, la GC es la técnica más ampliamente usada, contando con la opción de derivatización previa para aquellos compuestos con mayor polaridad. Por otro lado, la LC es aplicada para compuestos con baja volatilidad, baja estabilidad térmica y una alta polaridad. Por su parte, la técnica de NMR puede considerarse universal en cuanto a los compuestos a determinar si son de carácter orgánico; así como por ejemplo los análisis basados en $^1\text{H-NMR}$ permiten el estudio de todos los compuestos que presentan protones en su estructura, siendo una de sus mayores desventajas su escasa sensibilidad respecto a otras técnicas.



A la hora de la identificación y cuantificación de analitos en alimentos y muestras biológicas, las técnicas cromatográficas pueden ir acopladas a diferentes detectores clásicos como el de fotodiodos en fila o bien de MS, siendo este último el más usado actualmente en su modalidad de espectrometría de masas en tándem (MS/MS).

Los analizadores basados en MS se pueden dividir en dos grandes grupos: espectrometría de masas de baja resolución (*low resolution mass spectrometry*, LRMS) y espectrometría de masas alta resolución (*high resolution mass spectrometry*, HRMS). Los analizadores tipo LRMS se caracterizan por presentar resultados óptimos en términos de robustez y sensibilidad, mientras que los instrumentos tipo HRMS cuentan con mayor capacidad para analizar un alto número de compuestos. Entre los tipos de analizadores más destacados hoy en día se encuentran los de triple cuadrupolo (*QqQ*), Orbitrap, el sistema híbrido cuadrupolo/tiempo de vuelo (*quadrupole/time of flight*, Q-TOF) y el sistema híbrido cuadrupolo/Orbitrap (*quadrupole/Orbitrap*, Q-Orbitrap) [30].

Para generar los iones en fase gaseosa, hay diferentes tipos de fuentes de ionización, como la de ionización electrónica (EI) y la ionización química (CI), utilizadas con frecuencia en GC-MS, mientras que la ionización por electronebulización (ESI), en modo positivo o negativo, y la ionización química a presión atmosférica (APCI) se emplean comúnmente en LC-MS [31].

Por otro lado, técnicas extensamente aplicadas para estudios no cuantitativos como la NMR (conocida con dicha finalidad como NMR cuantitativa, *qNMR*), han demostrado en los últimos años una excelente capacidad en la determinación de la concentración de analitos

CAPÍTULO PRIMERO. MARCO TEÓRICO

mayoritarios en gran variedad de matrices complejas como productos naturales [32], suplementos dietéticos [33] o fármacos [34].

Del mismo modo, el desarrollo de herramientas informáticas que faciliten el procesado y la gestión de todos los datos obtenidos mediante este tipo de técnicas son una pieza clave para la interpretación fiable y útil de los mismos. El uso de análisis estadísticos multivariantes en los estudios metabolómicos es la sistemática más extendida para el manejo de la información adquirida, permitiendo la diferenciación entre distintos grupos y discriminación de los potenciales marcadores de las matrices objetivo [35].

Teniendo en cuenta todas las consideraciones descritas, el enfoque metabolómico debe realizarse con un cuidadoso diseño de cada una de las etapas que lo conforman, desde la selección y toma de muestras, almacenamiento y conservación, procedimiento de extracción/purificación, análisis, procesado de los datos, análisis multivariante, validación de los modelos estadísticos y la interpretación de los resultados que puedan extraerse del estudio.

Un estricto control de la seguridad agroalimentaria a través de los nuevos recursos analíticos puede proporcionar ventajas y beneficios directos para la salud humana; menor desperdicio de alimentos y aumento en la calidad y durabilidad de los productos. Además, el correcto control y seguimiento de los procesos de producción pueden verse optimizados dando lugar a beneficios indirectos como la disminución de uso de energía, así como la reducción de la huella de carbono resultante.

La presente Tesis Doctoral se ha centrado en la aplicación de herramientas de la metabolómica para la clasificación y/o identificación de marcadores químicos de calidad en bebidas alcohólicas mediante



técnicas de cromatografía-MS y NMR, a fin de aportar al control de calidad de ron dorado. Además, se ha llevado a cabo la aplicación de la técnica de NMR cuantitativa (*q*-NMR) para la determinación de distintas sustancias activas (ácido etidróico y ácido bórico). En las siguientes secciones se expondrán en detalle los fundamentos y la investigación desarrollada en la presente Tesis para cada uno de los apartados que la componen.

1.2 Metabolómica

1.2.1 Introducción al análisis metabolómico

El desarrollo de estudios metabolómicos permite el estudio de sistemas biológicos mediante la observación de las moléculas interconectadas a través de rutas metabólicas. El amplio potencial de este tipo de análisis, de la mano de las capacidades que aportan los avances tecnológicos, da lugar a estudios cada vez más exhaustivos y sofisticados de todo tipo de matrices biológicas y agroalimentarias.

La información que permite obtener el análisis metabolómico ayuda a entender las matrices en base al estudio de los analitos de bajo peso molecular presentes en ellas. Distintos factores intrínsecos (edad, maduración, etc.) o extrínsecos (sustancias añadidas, contaminantes, etc.) aportan información clave para conocer el estado del sistema en estudio y cómo se ve afectado por el entorno. Las herramientas de la metabolómica pueden ser aplicadas a casi todo tipo de matrices y enfocarse a multitud de casos de estudio.

CAPÍTULO PRIMERO. MARCO TEÓRICO

A tal fin, se abarcan diferentes estrategias cuya naturaleza depende del objetivo del estudio y el conocimiento previo del problema objetivo. A la hora de realizar un estudio metabolómico, es necesario un diseño previo, en el que hay que considerar diferentes etapas. Las etapas comunes de cualquier estudio metabolómico pueden resumirse en las siguientes:

- Establecimiento de la hipótesis de trabajo
- Recolección de muestras representativas
- Elección de la técnica o técnicas analíticas más adecuadas al estudio propuesto
- Pretratamiento de las muestras (p. ej. extracción, purificación o limpieza, preconcentración, derivatización, etc.)
- Análisis de las muestras, adquiriendo la máxima información útil posible
- Pretratamiento de los datos para su adecuación previa a los análisis quimiométricos
- Procesado de datos: Análisis multivariantes, tanto supervisados como no supervisados
- Validación de los modelos propuestos
- Identificación de marcadores con capacidad discriminante
- Interpretación de los resultados

En el siguiente esquema (Figura 1) se muestran los principales apartados de un estudio metabolómico tras la adquisición de los datos.

Un buen planteamiento del flujo de trabajo puede ser clave para el éxito del análisis metabolómico. Por ello, es de vital importancia la definición adecuada de los objetivos que se quieren alcanzar para así responder aquellas preguntas que se plantean.



La selección de las muestras ha de ser minuciosa, de forma que éstas sean representativas y aporten la información necesaria que se requiera obtener del estudio. Su correcta conservación y manipulación, evitando así cualquier tipo de degradación o pérdida de información, es esencial en este tipo de trabajos.

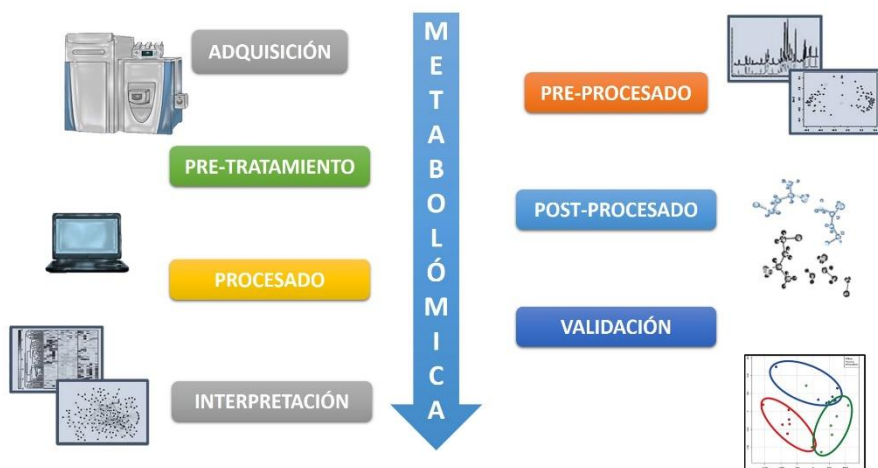


Figura 1. Flujo de trabajo general en un estudio metabolómico

La información previa del problema propuesto, así como de la información que se pretenda obtener, permitirá una correcta selección del tipo de análisis metabolómico que se precise. En aquellos casos donde el número y analitos de interés sean conocidos se aplicará el análisis dirigido (*targeted*). Un ejemplo de análisis dirigido puede ser la obtención de un perfil metabolómico, donde el enfoque se centra en el estudio de un conjunto de metabolitos específicos, bien por similitudes físico-químicas o por participar en una ruta metabólica concreta. Sin embargo, cuando se desconocen los metabolitos de interés que pueden tener influencia en el estudio, el análisis no dirigido (*untargeted*) es el

más adecuado. El objetivo en tal caso es disponer de la mayor información posible de la muestra sin precisar de la identificación de analitos concretos (huella metabolómica) [5,6].

En los estudios no dirigidos son clave tanto la extracción como el análisis a fin de obtener el mayor número de metabolitos posible para tener una información global de la problemática. A tal fin es necesaria la aplicación de técnicas de tratamiento de muestra y de análisis adecuadas. La selección de la técnica de análisis dependerá de las características del problema en estudio, siendo las técnicas más ampliamente utilizadas hoy en día con dicha finalidad la NMR y la MS acoplada a GC o LC [36].

El pretratamiento y tratamiento de los datos obtenidos de los análisis, sobre todo en los no dirigidos, han de ser cuidadosamente optimizados. La información, tanto de carácter cuantitativo como cualitativo es *a priori* desconocida, por lo que los datos obtenidos pueden llegar a ser de alta complejidad y volumen. Es por tanto necesario para un correcto tratamiento de éstos el uso de herramientas especializadas en análisis de datos que adapten la información obtenida a los análisis quimiométricos a realizar. La comparación y estudio de todos los resultados obtenidos permite la clasificación de muestras en categorías predefinidas, así como la identificación de posibles analitos característicos de una condición específica.

Finalmente, una estricta validación del modelo permitirá una óptima interpretación de los resultados. Para ello, será necesaria la división de datos en conjuntos de entrenamiento y de validación (p. ej. validación cruzada).

De forma adicional, una recopilación de datos procedentes de distintas técnicas de análisis en un único modelo puede conducir a una mejora en



las predicciones, así como a una mejora de la interpretación de los resultados. Este proceso de unión de información es la denominada fusión de datos.

En los siguientes apartados de la presente Tesis se describirán en profundidad las distintas etapas de las que consta un estudio metabolómico. En la Tabla 2 se reúnen a modo de resumen los términos más frecuentemente empleados en este tipo de estudios.

Tabla 2. Términos y definiciones comunes en la quimiometría aplicada a la metabolómica

Término	Definición	Referencias
Análisis dirigido (<i>targeted</i>)	Análisis centrados en un grupo específico de analitos previstos, los cuales generalmente requieren de la identificación y cuantificación de los mismos	[37]
Análisis discriminante	Análisis estadístico para la determinación de diferencias o semejanzas entre un grupo de variables predefinidas	[38]
Análisis multivariante	Análisis estadístico de más de una variable (p. ej. analitos) de forma simultánea	[36,39]
Análisis no dirigido (<i>untargeted</i>)	Análisis enfocado en la detección de tantos grupos de analitos como sea posible para obtener patrones o huellas dactilares sin necesariamente identificar o cuantificar uno o más compuestos específicos	[37]
Análisis predictivo	Extracción de información de un modelo con datos predefinidos para la predicción del comportamiento futuro o estimación de resultados desconocidos para cierta variable o variables	[40]

Término	Definición	Referencias
Análisis quimiométrico	Análisis que utiliza métodos matemáticos o estadísticos para diseñar o seleccionar procedimientos y experimentos de medición óptimos, proporcionando la máxima información química relevante mediante el análisis de datos químicos	[41]
Carga (<i>Loading</i>)	Influencia de las variables en un modelo multivariante	[42]
Componente principal	Combinación de variables originales que forman ejes de máxima dispersión de las muestras en el espacio del modelo	[42]
Modelo estadístico	Análisis matemático para el estudio de la correlación entre variables, generalmente datos X frente a datos Y (p. ej, PCA, PLS, LDA).	[43,44]
Modelo no supervisado	Determinan los conjuntos de patrones según sus propiedades y sin un conocimiento previo de la categoría de las variables (p. ej., PCA)	[45]
Modelo supervisado	Usan herramientas estadísticas con información adicional sobre los conjuntos de datos para determinar similitudes y diferencias entre grupos predefinidos (p. ej., LDA)	[45]
Puntuación (<i>Score</i>)	Para una observación dada, es la distancia desde el origen a lo largo de su componente hasta el punto donde se proyecta en el modelo	[42]
Validación estadística	Conjunto de evaluaciones que tienen como objetivo asegurar la calidad de las conclusiones extraídas del análisis de datos (p. ej. validación cruzada)	[44]





1.2.2 Adquisición de datos

Como antes se ha mencionado, en función del planteamiento inicial propuesto para el estudio metabolómico, la información que se quiera obtener, así como el conocimiento previo que se tenga de la matriz objetivo, pueden abordarse diferentes aproximaciones.

Un análisis dirigido (*targeted*) está orientado a la determinación (cuantitativa o cualitativa) de una serie de compuestos relacionados por alguna propiedad o característica en lugar de estudiar el metaboloma completo. La ventaja de este tipo de análisis reside en simplificar el manejo de datos obtenidos al centrarse en ciertos marcadores concretos. Por otro lado, el análisis no dirigido (*untargeted*) se centra en la detección del mayor número posible de analitos, con el fin de asociar patrones o relaciones específicas de ellos respecto a una condición, característica o tratamiento. La selección tanto de los métodos de extracción como de la técnica de análisis dependerá del tipo de enfoque seleccionado para el estudio [13].

La etapa de extracción de compuestos volátiles/semivolátiles de las matrices puede miniaturizarse y automatizarse con el uso de técnicas como la microextracción en fase sólida (SPME) (con posible derivatización previa), entre otras, minimizando los errores y contaminaciones debidos a la manipulación de la muestra. Mientras que para los componentes no volátiles se encuentra muy extendido el empleo de métodos de tratamientos genéricos como “*dilute and shoot*” (diluir e inyectar), que permiten una extracción rápida, repetitiva y reproducible de las muestras sin perder información discriminante [30].

Entre las técnicas de análisis más empleadas para estudios metabolómicos se encuentra la cromatografía, debido al alto grado de eficacia en la separación de analitos objeto de estudio, facilitando así el estudio individualizado de cada uno de ellos y la capacidad de obtener amplia información en cuanto a número de analitos en todo tipo de matrices. Dependiendo de las características físico-químicas (polaridad, volatilidad, etc.) de los compuestos que se pretenden determinar, será la GC o LC la técnica óptima para la separación de los mismos, siendo muchos compuestos susceptibles a ambos tipos de análisis. No obstante hay que mencionar que otras técnicas también son ampliamente empleadas a tales fines, destacando entre ellas por la cromatografía iónica (IC) [46], ICP-MS [47] o la espectroscopía de absorción atómica (EAA) [47].

Como ya se ha expresado, las técnicas tanto de GC y LC pueden acoplarse a analizadores de MS de baja resolución, como QqQ o trampa de iones, o de alta resolución como sector magnético o analizadores de tiempo de vuelo (TOF). En el caso de los análisis no dirigidos, la adquisición por MS en modo barrido completo (*full scan*) es la más utilizada, obteniéndose así las medidas de masa exacta de la mayor cantidad de analitos presentes. Para análisis dirigido, los estudios requieren de una identificación estricta. A tal fin, el modo de monitorización de reacción múltiple (MRM), donde los iones precursores dan lugar a fragmentos específicos (iones producto), es el más usado.

La NMR de alta resolución también se ha aplicado en estudios metabolómicos. A pesar de su alto coste, esta técnica, presenta importantes ventajas ante la posibilidad de analizar cantidades



relativamente pequeñas de muestra o extractos de muestra de una manera no destructiva, fácil y directa, permitiendo la determinación simultánea de varios analitos en un mismo análisis.

Aunque la aplicación de la NMR se extiende a un apreciable número de núcleos magnéticamente activos (hidrógeno-1, carbono-13, nitrógeno-15, flúor-19, fósforo-31, etc.), el procedimiento más extendido en la metabolómica es el monodimensional de hidrógeno ($^1\text{H-NMR}$). Éste hecho se debe tanto a su presencia en prácticamente la totalidad de los analitos presentes en las muestras, junto a su elevada abundancia natural y sensibilidad [23].

En matrices que contienen altas concentraciones de componentes mayoritarios, como es el caso del agua y el etanol en bebidas alcohólicas, puede perderse información de aquellos minoritarios en adquisiciones mediante NMR. La distorsión generada en los espectros obtenidos debido a la alta intensidad de las señales puede evitarse con pulsos selectivos a distintas frecuencias para la supresión de éstas. De éste modo se logra el análisis y estudio de señales solapadas y con alto poder discriminante en los estudios metabolómicos de bebidas con elevado contenido alcohólico, como son las bebidas espirituosas [48].

1.2.3 Pretratamiento de señales instrumentales

Los efectos de errores tanto en el proceso de preparación de la muestra como en la medida de la señal analítica pueden verse reducidos con un adecuado pretratamiento de datos. Dicho pretratamiento consta de

distintas etapas, siendo la mayoría de ellas comunes para la información de MS o de NMR.

El alineamiento de cromatogramas y espectros es una de las etapas más importantes. Permite comparar las señales obtenidas pese a los posibles desplazamientos experimentales, tanto en tiempo de retención (RT) como en el desplazamiento químico (δ). Dichas diferencias pueden ser debidas a interacciones de los componentes de las matrices de estudio o a causas experimentales como ligeras desviaciones en la adquisición de las mismas.

En la NMR, la corrección de la línea de base es otra técnica de pretratamiento espectral que es crítica para eliminar las señales que pueden surgir de distorsiones electrónicas, filtrado digital inadecuado o muestreo digital incompleto. Cuando se realiza correctamente, la corrección de la línea base produce un espectro de aspecto más fidedigno donde las regiones sin señal son líneas horizontales completamente planas con intensidad cero [49].

En el caso de cromatogramas obtenidos mediante GC y LC, la deconvolución de la señal obtenida permite resolver el solapamiento de los picos, generando una tabla donde cada variable corresponderá a una señal resuelta. Este paso puede reducir considerablemente la complejidad de los datos obtenidos. Hoy en día existen diversos *software* para la deconvolución orientada a la MS basados en el estudio de los perfiles espectrales de las señales para la resolución de los picos (Tabla 3).

En el caso de la NMR, el estudio de las señales está basado en la segmentación de los espectros obtenidos, logrando así una comparación regional entre muestras. A diferencia de la información obtenida



mediante LC-MS o GC-MS, donde la señal correspondiente a uno o varios analitos aparece a un RT, en los espectros de NMR las señales a diferentes δ pueden ser debidas al mismo analito, con sus correspondientes desdoblamientos magnéticos. Es por ello que la segmentación de los espectros para comparaciones regionales equitativas es la técnica más extendida en los estudios metabolómicos mediante NMR, reduciendo la dimensionalidad de los datos a costa de la pérdida de resolución espectral útil para posibles futuras identificaciones. La segmentación implica la división de los espectros de NMR en pequeñas regiones (generalmente de 0.01–0.1 ppm), que son lo suficientemente anchas como para incluir una o más señales de NMR. La segmentación de la NMR da lugar a un conjunto de datos de intensidades resultantes de la fragmentación equidistante de los espectros. Regiones de compuestos comunes y mayoritarios, como el caso del agua en la mayoría de matrices, pueden excluirse selectivamente de la segmentación [58].

Tabla 3. Listado de software disponible para deconvolución en MS

Término	Accesibilidad	Referencias
AnalyzerPro	Disponible comercialmente	[50]
Automated Mass Spectrometry Deconvolution and Identification (AMDI)	Gratuito	[51]
ChromaTOF	Disponible comercialmente	[52]
Deconvolution Reporting Software (DRS)	Disponible comercialmente	[53]
MetaboliteDetector	Gratuito	[54]
MS-DIAL	Gratuito	[55]
MzMine	Gratuito	[56]
TraceFinder	Disponible comercialmente	[57]

La disponibilidad de *software* orientado al alineamiento de espectros de NMR es más reducida que en el caso de la MS, siendo programas como NMRProcFlow, MestreNova o Chenomx NMR Suite los más comúnmente utilizados [49,59].

El escalado es otra parte importante del pretratamiento de datos, tanto de los obtenidos mediante MS como NMR. Cuando se trabaja con datos para estudios no dirigidos, un análisis directo de los datos sin tratamiento daría lugar a análisis con enmascaramiento de variaciones en la matriz al usarse valores de variables basadas en el área obtenida. Por ello, la aplicación de estrategias de escalado de datos es prioritaria antes de cualquier estudio metabolómico, centrándose así en la variación de la composición de la muestra y en las similitudes/diferencias entre el conjunto de datos. Los tipos de escalado más comúnmente empleados en metabolómica se muestran a continuación desde la Ecuación 1 a la Ecuación 6 [49]:

- **Centrado medio:** Este método sustrae la media ($\overline{x_{ij}}$) a cada una de las medidas de una variable (x_{ij}), de manera que solo tenemos la variabilidad presente entre las muestras:

$$x_{ij} = x_{ij} - \overline{x_{ij}}$$

Ecuación 1. Método de escalado mediante centrado medio

- **Autoescalado (escalado unitario):** El objetivo del autoescalado es dar el mismo peso a todas las variables (centrado medio dividido por la desviación estándar de cada variable, s_i). De este modo los analitos con baja y alta abundancia contribuirán de igual forma en los modelos multivariantes. Por otro lado, el



autoescalado destacará aquellas señales poco relevantes para el estudio, incrementando además los errores de medición de aquellas variables de poco interés que puedan aparecer. Por ello, su uso requiere de un conocimiento previo de la matriz de estudio, suprimiendo aquellas variables que puedan interferir.

$$x_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{S_i}$$

Ecuación 2. Método de transformación de escalado unitario (autoescalado)

- **Escalado Pareto:** Similar al autoescalado, pero en este caso cada elemento en la característica centrada a la media se divide por la raíz cuadrada de la desviación estándar. El escalado Pareto es un compromiso entre el centrado medio y el autoescalado, donde las variables con alta abundancia son menos dominantes en comparación con los correspondientes centrados a la media. Es el modo de escalado más ampliamente usado para datos obtenidos mediante detectores tipo MS y NMR, restando la importancia del ruido mientras preserva la varianza del conjunto de datos.

$$x_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{\sqrt{S_i}}$$

Ecuación 3. Método de transformación de escalado Pareto

- **Escalado por rango:** Cada variable ya centrada a la media se divide por el rango. Esta operación mejora la importancia en el modelo de aquellas variables con menor variabilidad a la vez que reduce notoriamente la varianza. No funciona bien cuando la diferencia entre el máximo y mínimo del rango es muy grande.

$$x_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{(x_{max} - x_{min})}$$

Ecuación 4. Método de transformación de escalado por rango

Por otro lado, la aplicación de métodos de transformación matemática para la corrección de la heterodasticidad es muy extendida en metabolómica. La transformación logarítmica y la transformación de poder son las más comunes [60]:

- **Transformación logarítmica:** Cada variable original se reemplaza por el valor de su logaritmo decimal. De esta forma los valores iniciales son reducidos a números próximos a cero, reduciendo el rango de las variables sin enmascarar el efecto de aquellas más pequeñas y evitando que los valores mayores tengan influencia sobre los menores.

$$x_{ij} = \log(x_{ij})$$

Ecuación 5. Método de transformación logarítmica

- **Transformación de potencia:** Cada variable original se reemplaza por el valor de su raíz n ; siendo n (Ecuación 6)



seleccionado de forma arbitraria. Es una buena alternativa cuando la transformación logarítmica no se puede conseguir por la presencia de ceros en las variables, los cuales dan resultados próximos a menos infinito en tales casos.

$$x_{ij} = \sqrt[n]{x_{ij}}$$

Ecuación 6. Método de transformación de potencia

Una vez que se han procesado todos los datos se obtiene una matriz compuesta por un número de observaciones (muestras, Y) y un número de variables (señales o analitos, X), preparada para un consecuente análisis multivariante. La Tabla 4 reúne las definiciones de los términos más comúnmente usados en el pretratamiento de señales en estudios metabolómicos.

Tabla 4. Términos y definiciones comunes en el pretratamiento de datos de estudios metabolómicos

Término	Definición	Referencias
Alineamiento	Sincronización del RT o δ de varios cromatogramas o espectros de NMR, respectivamente, en un conjunto de muestras dado	[44]
Corrección de línea base	Método para la rectificación de inclinaciones y desviaciones de los espectros de NMR	[59]
Deconvolución	Resolución de picos superpuestos en un cromatograma, generalmente LC o GC, utilizando una segunda dimensión (como la MS)	[44]

Término	Definición	Referencias
Eliminación de valores atípicos	Eliminación de valores que exhiben una alta desviación respecto la mayoría de sus réplicas	[44]
Escalado	Reducción del impacto de señales en modelos con gran varianza, destacando aquellas que presentan mayor variabilidad, pero menor concentración	[58]
Normalización	Corrección de la variabilidad existente entre las distintas muestras del estudio debido a variaciones experimentales	[58]
Rellenado de ceros	Adición de valores a aquellos datos nulos en la tabla obtenida tras la selección de señales	[41]
Segmentación	Método de definición de tamaño de segmentos espectrales para su integración individual. Generalmente aplicado para señales de NMR	[41]
Selección de variables	Estudio de señales significativas de GC-MS o LC-MS para la obtención de tablas con valores m/z correspondientes a intensidades de señales	[44]
Suavizado (<i>Smoothing</i>)	Optimización de la resolución del cromatograma o espectro obtenido para una mejor robustez del análisis e interpretación del mismo	[36]
Transformación matemática	Alteración matemática para la corrección de datos sesgados que puedan ser originados por situaciones biológicas particulares (heterocedasticidad)	[35]





1.2.4 Métodos multivariantes

Se pueden diferenciar dos tipos de métodos de análisis estadísticos multivariantes, supervisados y no supervisados, en función de si se conoce e indica la pertenencia de los datos Y a distintas categorías o no. Generalmente se emplean en primer lugar los métodos estadísticos no supervisados para una exploración inicial del conjunto de datos. A tal fin, existen diversos *software* destinados a la aplicación de éste tipo de análisis en estudios metabolómicos, En la Tabla 5 se muestran algunas de las herramientas actualmente más extendidas en éste tipo de estudios.

Tabla 5. Herramientas para el análisis estadístico de datos en estudios metabolómicos

Término	Accesibilidad	Referencias
Compound Discoverer	Disponible comercialmente	[61]
MATLAB	Disponible comercialmente	[62]
MetaboAnalyst	Gratuito	[63]
MetabolAnalyze	Gratuito (Complemento de R)	[64]
MZmine	Gratuito	[56]
SIMCA Umetrics	Disponible comercialmente	[65]
Unscrambler	Disponible comercialmente	[66]
Workflow4metabolomics	Gratuito	[67]
XCMS Online	Gratuito	[68]

Los métodos no supervisados muestran las distintas relaciones existentes entre las diferentes X variables e Y observaciones, siendo así observables las similitudes y diferencias de forma global. La optimización de los parámetros de pretratamiento de datos previamente descrita puede llevarse a cabo con este tipo de análisis. La interpretación de los resultados permitirá además la detección de muestras atípicas, *outliers*. Dichas observaciones muy diferentes a las demás serán eliminadas,

obteniendo así un conjunto de muestras válidas para el estudio, con X variables con una desviación apropiada.

Generalmente, el análisis estadístico no supervisado más utilizado en los estudios metabolómicos es el análisis de componentes principales (PCA). Este tipo de análisis ofrece una interpretación del conjunto de datos, sin información previa a ninguna categoría, explicando la variabilidad de la información con un número relativamente pequeño de componentes principales (PCs) [69].

La selección del número óptimo de PCs es crucial para la interpretación de los resultados del PCA: una sobreestimación o infraestimación del modelo puede dar importancia a variables como el ruido, así como puede además omitir información relevante en los resultados, respectivamente. Uno de los métodos más empleados para conseguir una buena estimación de PCs es la validación cruzada (CV). Este método calcula el error de predicción que se consigue al elegir diferente número de variables para cierto número de PCs. Es decir, después de ordenar las nuevas variables o PCs en función a su influencia en el modelo, éstas mismas se van agregando secuencialmente hasta incluir todas, registrando el error obtenido en cada modelo y estableciendo así el número óptimo. Dicho error es evaluado por el valor de ajuste del modelo (R^2) y por su capacidad de predicción (Q^2). El resultado óptimo será aquél que, tras la adición de cierto número de PCs, aporte un aumento ínfimo al error obtenido [35].

Otro tipo de análisis no supervisado comúnmente usado en los estudios metabolómicos para la reducción de la dimensionalidad de los modelos en interpretaciones bidimensionales es el análisis de agrupamiento jerárquico (HCA), en el que se enfatizan los grupos y patrones naturales



en un diagrama ascendente. La agrupación final se logra mediante el uso de una métrica apropiada en la distancia de las muestras, organizando los resultados en dicha jerarquía gráfica. El HCA permite, además, evaluar la relación entre distintos grupos de observaciones y variables mediante interpretación de sus distancias [70].

Normalmente, a continuación de los análisis no supervisados, se procede a la aplicación de análisis supervisados, donde las variables iniciales de los modelos son descritas en función a sus categorías de pertenencia, para una interpretación focalizada en sus agrupaciones. Gracias a estos análisis es posible obtener modelos de clasificación a partir de las variables pertenecientes a las observaciones analizadas. A tal fin, se utiliza un conjunto mayor de muestras para el perfeccionamiento del modelo (conjunto de entrenamiento), mientras que el resto se usa para confirmación en la predicción de las mismas (conjunto de predicción).

Entre los diversos tipos de análisis supervisados ampliamente utilizados en estudios metabolómicos se pueden citar, el análisis discriminante lineal (LDA), el análisis discriminante por mínimos cuadrados parciales (PLS-DA) o el análisis discriminante por mínimos cuadrados parciales ortogonal (OPLS-DA). La Tabla 6 describe distintos tipos de análisis y términos comúnmente utilizados en modelos multivariantes.

Tabla 6. Denominación de análisis y términos en modelos multivariantes

Término	Definición	Referencias
AUC	Valor entre 0 y 1 correspondiente al área bajo la curva ROC. Cuando el valor es próximo a 1, se considera que la capacidad clasificatoria del modelo es óptima	[35]
CV-ANOVA	Análisis de varianza con validación cruzada que proporciona un valor p para el cálculo del nivel de separación de grupos en los análisis OPLS	[71]
Datos Y	Variables con valores cuantitativos o cualitativos relacionados con las observaciones del modelo (p. ej. señal analítica)	[39]
Datos X	Número de observaciones de un modelo estadístico (p. ej. muestras biológicas)	[39]
DModX	Distancia de una observación dada al modelo para la detección de posibles valores atípicos, relacionado con la suma de cuadrados residual o error cuadrático de predicción	[71]
HCA	Análisis no supervisado para la búsqueda de agrupaciones por similitud tanto de observaciones como variables en el modelo	[41]
Hotellings T^2	Generalización de la distribución de las variables en un intervalo de confianza del 95% que puede ser usada para identificar valores atípicos en un modelo estadístico	[71]
LDA	Análisis supervisado para maximizar la varianza entre clases disminuyendo la varianza dentro de las mismas a través de funciones lineales	[35]
OPLS-DA	Extensión del análisis PLS-DA donde la varianza entre los grupos propuestos no se encuentra relacionada entre sí, dando mayor importancia a los componentes predictivos	[71]
PCA	Análisis no supervisado para la reducción de dimensiones del modelo en componentes principales mediante transformaciones lineales, manteniendo la mayor varianza posible	[39]





Término	Definición	Referencias
PLS-DA	Análisis discriminante supervisado de las variables de un modelo para la predicción de observaciones relativas a varias clases mediante reducción de dimensiones	[72]
Q²	Medida de la capacidad del modelo para predecir un nuevo conjunto de datos, explicada mediante validación cruzada de los datos	[71]
R²	Medida de los datos originales correctamente explicados por el modelo propuesto (ajuste del modelo), siendo el valor igual a 1 cuando se da un valor de acierto del 100% de los casos	[71]
ROC	Análisis para la evaluación de la habilidad predictiva de un modelo por representación gráfica de sensibilidad frente a especificidad, generalmente en clasificaciones binarias	[35,41]
VIP	Análisis de la carga discriminatoria de las variables en un modelo clasificatorio, dando valores superiores a 1 cuando la contribución es mayor que la media	[71]

Tanto el análisis LDA como PLS-DA son hoy en día los análisis supervisados más ampliamente utilizados en estudios metabolómicos tanto de carácter sanitario como agroalimentario [73]. El primero se basa en la maximización de la relación de varianza entre clases y la relación dentro de las mismas, usando así combinaciones lineales de las variables originales con el fin de lograr modelos de clasificación. Al requerir el LDA de un número semejante de muestras y de variables, es común la selección de modelos donde primero se produce una etapa de extracción de variables, como es el caso de los PLS-DA [74].

Aunque el PLS-DA implica una serie de operaciones matemáticas y una gran cantidad de parámetros, fundamentalmente su aplicación conlleva

dos procesos principales; (a) reducción de la dimensionalidad mediante construcción de componentes; y (b), la construcción del modelo discriminante. Mediante métodos de regresión iterativa se construyen modelos que relacionan un grupo de variables independientes y una variable dependiente, permitiendo así maximizar la covarianza entre los grupos. La representación gráfica de estos modelos, de manera similar al PCA, es útil para interpretar el grado de discriminación entre los grupos, con una mejora sustancial en términos de clasificación [72].

Una modificación del PLS-DA es el OPLS-DA, similar a su predecesor, donde las variaciones de la matriz de datos son divididas en dos partes a través de una corrección de señal ortogonal: una parte muestra la capacidad de respuesta lineal, mientras que la otra muestra una respuesta ortogonal a la lineal. Aquellas variables que muestren una variación considerable con la respuesta resultante serán usadas para la creación del modelo, consiguiendo así una maximización entre los grupos correspondientes a las categorías iniciales [35].

1.2.5 Validación de modelos multivariantes

Tras el diseño y construcción de los modelos estadísticos, es imprescindible llevar a cabo un proceso de validación para demostrar la aptitud de los mismos. La validación del modelo es un proceso para la evaluación de los resultados obtenidos, garantizando que las relaciones hipotéticas entre las observaciones/muestras y las variables/respuestas proporcionan una capacidad óptima de predicción del modelo. Modelos de predicción con PLS-DA tienen una alta tendencia al sobreajuste, por lo que es necesaria una validación cuidadosa.



Pese al apreciable incremento de estudios metabolómicos en los últimos años, así como al desarrollo cada día de herramientas más completas para la aplicación de análisis multivariantes, la existencia de guías predefinidas para abordar la validación de modelos metabolómicos es muy reducida. alguna de ellas, como la guía para la validación de análisis multivariantes de la Federación Europea de Asociaciones Nacionales de Laboratorios de Medición, Ensayo y Análisis (Eurolab), tiene como finalidad la evaluación de estudios metabolómicos basados en la NMR [75]. En la mayoría de los casos, la validación se ve reducida a aquella incluida por los desarrolladores del *software*, comúnmente basadas en CV.

La forma más irrefutable para la evaluación de la capacidad predictiva del modelo pasa por la evaluación de nuevas observaciones de un conjunto de datos independiente: el ya previamente descrito conjunto de predicción. Dichas observaciones, generalmente un 20%-25% de las disponibles, no son usadas en la construcción del modelo. El coste de dicha división es la posible pérdida de información sustancial para el estudio metabolómico, por lo que sólo debe de realizarse cuando hay suficientes observaciones disponibles para no afectar a los resultados finales. La evaluación del conjunto de predicción proporcionará un valor inequívoco de muestras correctamente clasificadas (CC%). Aparte de la evaluación mediante conjuntos de entrenamiento y predicción, otra de las estrategias más comúnmente utilizadas es la CV.

Los resultados del procedimiento de CV se basan en la optimización en torno a parámetros de calidad, como R^2 y Q^2 . R^2 mide la bondad del ajuste del modelo al conjunto de datos usados, mientras que Q^2 mide la

capacidad predictiva del mismo. Un valor de R^2 igual a 1 indica un ajuste completo de los datos por el modelo, mientras que Q^2 igual a 1 indica una capacidad predictiva del 100% de acierto (generalmente $Q^2 > 0.5$ se admite como un buen valor de predictibilidad). El valor R^2 aumenta a medida que aumenta el número de PCs, aproximándose a 1 si cuenta con un número equiparable a la matriz X . Por otro lado, el valor de Q^2 puede sufrir una disminución con el aumento de PCs debido al ajuste en base a variables del conjunto de datos que ya no son representativas de la población estudiada. Es decir, en cierto grado de complejidad, la capacidad de predicción del modelo disminuye. Una gran discrepancia entre R^2 y Q^2 indica un sobreajuste del modelo mediante el uso de demasiados componentes [76].

Pueden diferenciarse varios tipos de CV (*Leave One Out*, Monte-Carlo, 2CV, etc.), siendo la más extensamente utilizada la validación cruzada dejando k -muestras fuera (*k-fold CV*). En ella, el conjunto de datos es dividido en k número de subconjuntos y, para un número fijo de PCs, todos los individuos de cada subconjunto son clasificados usando un submodelo construido con los otros $k-1$ subconjuntos (subconjunto de calibración). Las diferencias entre los valores pronosticados y los valores son entonces utilizadas para calcular el parámetro Q^2 para este número de PC. El procedimiento comienza en $PC = 1$ y se repite incrementando el valor siempre que el aumento de Q^2 sea mayor que un valor óptimo fijado [76].

Otra estrategia de validación de un modelo clasificadorio muy extendida es el análisis de característica de funcionamiento del receptor (ROC). Un gráfico ROC presenta la habilidad de acierto en el eje de abscisas frente al ratio de falsos positivos obtenidos ($100 - \text{especificidad}$), en el eje de



ordenadas, formando la conocida como área bajo la curva (AUC). El valor de AUC puede aplicarse entonces como prueba de capacidad clasificatoria, asignando un valor igual a 1 a una prueba perfecta, mientras que un valor de AUC de 0.5 indicará que la mitad de los resultados son debidos al azar. Si bien el análisis ROC ofrece la posibilidad de ajustar el valor de corte para optimizar las estrategias de clasificación, su uso siempre es recomendable como análisis adicional al resto de validaciones [77].

Además, los modelos multivariantes construidos pueden evaluarse a través de distintos tipos de errores: [78]:

- Raíz del error cuadrático medio de calibración (RMSEC): ajuste del modelo al conjunto total de datos del estudio metabolómico.
- Raíz del error cuadrático medio de la validación cruzada (RMSECV): medida de la capacidad del modelo para la predicción de observaciones que no fueron empleadas para la construcción del modelo mediante el proceso de validación cruzada.
- Raíz del error cuadrático medio de predicción (RMSEP): aporta información acerca del ajuste del modelo a aquellos datos y variables que no fueron empleadas para la construcción del modelo.

Todos los valores de las distintas raíces de error cuadrático medio (RMSE) pueden ser determinadas mediante la Ecuación 7, donde el valor y es el valor real, mientras que el valor \hat{y} es el valor predicho, el cual difiere entre los distintos tipos de RMSE. Un valor bajo de los distintos

RMSE es indicativo de un buen comportamiento del modelo desarrollado [79]:

$$RMSE = \sqrt{\frac{1}{N} \sum (y - \hat{y})^2}$$

Ecuación 7. Ecuación para el cálculo de la raíz del error cuadrático medio

Otros valores adicionales para la validación de un modelo multivariante, como la sensibilidad y la selectividad, pueden ser calculados mediante la aplicación de los resultados de clasificación obtenidos [75]:

- Sensibilidad: Positivos verdaderos / (positivos verdaderos + falsos negativos)
- Selectividad: Negativos verdaderos / (negativos verdaderos + falsos negativos)

Finalmente, puede llevarse a cabo la evaluación de la estabilidad del modelo para garantizar su aplicabilidad en distintas condiciones. Por ejemplo, los resultados de la medición están sujetos a la variación del instrumento debida a operaciones de mantenimiento o de las condiciones ambientales, lo que puede afectar la estabilidad a largo plazo del modelo. Una evaluación del sistema podría realizarse aplicando nuevas muestras medidas después de un período de tiempo considerable, por otro analista, instrumento o laboratorio (validación inter-laboratorio).

1.2.6 Marcadores metabólicos

Tras la validación de los modelos multivariantes, las señales más discriminantes pueden estudiarse a fin de identificar potenciales



marcadores de las matrices de estudio. El objetivo es poder correlacionar dichos marcadores con las distintas propiedades de las mismas.

Los resultados de los análisis supervisados de los modelos multivariantes permiten la medición de capacidades discriminantes de las variables del conjunto de datos. Entre las técnicas más empleadas hoy en día para la evaluación de marcadores indicativos se encuentran los gráficos *S-plot*, los análisis Jack-Knife y el análisis de la importancia de la variable en la proyección (VIP) [41].

Tanto los gráficos *S-plot* como el análisis Jack-Knife aportan información orientada a la detección de agrupaciones de variables discriminantes. Por un lado, *S-Plot* muestra la covarianza y correlación entre las distintas variables a través de un diagrama de dispersión. Ello permite la determinación grupal de las mismas en forma de una gráfica de puntuaciones. Por otro lado, Jack-Knife adopta la forma de un diagrama de columnas, indicando la incertidumbre y estabilidad de las variables y permitiendo la extracción de grupos de marcadores con alto poder de discriminación.

Sin embargo, la aproximación más extendida para la evaluación discriminante de marcadores en los estudios metabolómicos es el análisis VIP. Este tipo de análisis se basa en la medición de la contribución de cada variable al modelo multivariante, seleccionándose normalmente aquellas con valor de puntuación $VIP > 1$ como variables indicativas [80].

La identificación de los analitos más discriminantes del estudio metabolómico en MS puede abordarse de diferentes formas. Si se trata de estudios realizados mediante información obtenida por GC-MS, el

análisis puede llevarse a cabo mediante comparaciones del espectro experimental con los teóricos obtenidos a partir de librerías espectrales comerciales y la coincidencia del RT con una biblioteca de espectros con RT fijo (p. ej. índice de retención de Kovats). En las librerías espectrales se compara la masa del valor de m/z obtenido experimentalmente en un espectrómetro de masas y las masas teóricas de los analitos contenidos en dichas bases de datos. En la Tabla 7 se reúne información sobre una serie de librerías comunes en GC y LC acopladas a MS para la identificación de marcadores en estudios metabolómicos, junto a sus características más destacables. Gracias a su reproducibilidad y la existencia de abundantes y extensas librerías de espectros de masas, la GC-MS está considerada la técnica de referencia para la identificación de marcadores [35].

Tabla 7. Bases de datos más utilizadas en la identificación tentativa de los analitos

Término	Descripción	Referencias
ChEBI	Base de datos elaborada por el Instituto Europeo de Bioinformática, perteneciente al Laboratorio Europeo de Biología Molecular (EMBL). Incluye más de 45.000 compuestos de distintas fuentes para proporcionar descripciones más estandarizadas de entidades moleculares	[81]
ChemSpider	Perteneciente a la <i>Royal Society of Chemistry</i> , consiste en una base de datos de compuestos orgánicos que proporciona un acceso rápido a más de 58 millones de estructuras, propiedades e información asociada	[82]
HMDB	Base de datos para la obtención de información derivada de metabolitos del cuerpo humano	[83]
KEGG	Se trata de la Enciclopedia de Genes y Genomas de Kyoto. Ofrece información	[84]





Término	Descripción	Referencias
	acerca de sustancias químicas y reactividad de compuestos relacionados con los seres vivos de cinco bases de datos distintas: <i>Compound</i> , <i>Glycan</i> , <i>Reaction</i> , <i>Rpair</i> , y <i>Enzyme</i>	
Librería NIST	Desarrollada por el Instituto Nacional de Normas y Tecnología de EEUU Contiene información acerca de compuestos orgánicos basados en análisis de técnicas como MS, infrarrojo medio o UV-Vis	[85]
LipidMaps	Base de datos enfocada a las sustancias de carácter lipídico, con información sobre su clasificación, nomenclatura y estructura	[86]
MassBank	Base de datos focalizada en el espectro de masas de alta resolución de compuestos	[87]
METLIN	Base de datos especializada en MS que ofrece información acerca de la fórmula molecular y estructura de analitos (espectros de masas y espectros de fragmentación). Elaborada por el Centro Scripps para la Espectrometría de Masas de Estados Unidos	[88]
PubChem	Contiene descripción e información estructural de millones de compuestos, la mayoría con peso molecular inferior a 2000 Da. Elaborada por el Centro Nacional de Información Biotecnológica de Estados Unidos	[89]

En cambio, en LC-MS la identificación de analitos a partir de librerías de MS no es adecuada debido a la variación de los ajustes experimentales, como las condiciones cromatográficas y los parámetros de adquisición, así como por la complejidad de los mecanismos de fragmentación. En general, las herramientas disponibles actualmente

para la identificación de datos de LC-MS se desarrollan en base a dos aspectos adicionales como la distribución isotópica y espectros de fragmentación obtenidos mediante MS/MS.

La interpretación de los perfiles isotópicos es de gran importancia debido a la capacidad que otorgan de discernir la molécula objetivo en base a la abundancia de aquellos isótopos naturales minoritarios presentes. En la actualidad, la mayoría de las herramientas informáticas de identificación para MS, cuentan con la capacidad de comparación de perfiles isotópicos teóricos con perfiles isotópicos experimentales que permiten un rango de candidatos para una identificación tentativa (p. ej. *Compound Discoverer*) [61].

Adicionalmente la información de la fragmentación de los compuestos obtenida mediante MS/MS puede usarse como criterio de confirmación. Los patrones de fragmentación que se obtienen en dicho modo de adquisición permiten su comparación con los obtenidos experimentalmente o con programas de simulación de fragmentación molecular. La inyección y análisis de patrones de aquellos analitos identificados tentativamente puede realizarse como confirmación conclusiva por comparación con los datos experimentales obtenidos en las observaciones originales del estudio metabolómico [3].

Por otro lado, la identificación de analitos por NMR en estudios metabolómicos generalmente se realiza en dos pasos. Primero, el espectro de NMR de la muestra se deconvoluciona en resonancias individuales o grupos de resonancias que pertenecen a un componente individual. En el segundo paso, estas huellas digitales espectrales se comparan con la información disponible en bases de datos de compuestos de NMR (p. ej. HMDB) [83]. El éxito de este enfoque para la



identificación positiva de un analito no solo depende de la calidad de la deconvolución espectral, sino que también requiere la presencia del espectro de NMR del compuesto en la base de datos, medido en condiciones similares o idénticas a las de la mezcla [90]. La alta complejidad que suelen presentar las matrices biológicas, generando solapamientos de señales espectrales, así como la dificultad de la correlación de distintas señales a lo largo del espectro, dan lugar a sistemáticas más tediosas para la identificación de marcadores mediante NMR. El acoplamiento para fraccionamientos controlados de fases móviles mediante LC, así como la adquisición de espectros bidimensionales (p. ej. ^1H - ^1H COSY) son las técnicas más extendidas [91,92].

1.2.7 Fusión de datos

Debido a la disponibilidad de múltiples conjuntos de datos obtenidos a partir del mismo sistema biológico, existe una creciente necesidad de analizar y visualizar dichos bloques de datos en conjunto para llegar a una mejor comprensión del sistema en estudio. La fusión de datos se define como la integración de bloques de datos de diferentes plataformas analíticas en un solo modelo. Por lo general, en los campos relacionados con la Química Analítica y las ciencias ómicas, la fusión de datos de múltiples plataformas tiene dos objetivos específicos: aumentar la capacidad de predicción y obtener una mejor comprensión de los fenómenos estudiados [93].

Actualmente las tres formas más extendidas de fusión de datos consisten en la fusión de nivel bajo, medio o alto, respectivamente [94].

- Fusión de nivel bajo: se basa en la concatenación básica de bloques de datos de diferentes técnicas analíticas después del preprocesado de los mismos.
- Fusión de nivel medio: combinación de variables tras una reducción de dimensionalidad mediante modelos estadísticos (p. ej. PLS-DA).
- Fusión de nivel alto: se obtiene mediante la combinación de resultados obtenidos por modelos previamente creados de forma independiente.

Adicionalmente, una cuarta forma de fusión de datos es la conocida como fusión de datos Kernel. Los métodos basados en este tipo de fusión emplean funciones matemáticas tipo Kernel para transformar datos en espacios de características de alta dimensión y generar matrices únicas. Dicho tipo de fusión de datos es menos usado en estudios metabólicos y se encuentra orientado a conjuntos de datos que presentan respuestas lineales [95].

La selección del tipo de fusión a aplicar depende de la diferencia entre los conjuntos de datos. Aquellos con diferencias más grandes a menudo implican niveles más altos de fusión de datos. La fusión de nivel alto implica el modelado y procesado de cada bloque de datos por separado, para evaluar los resultados de cada modelo conjuntamente. Además del mayor consumo de tiempo de éste modo de fusión, los resultados están sujetos a una pérdida de información considerable debido a que no se estudian las posibles relaciones entre las distintas variables. Su uso se centra en situaciones donde el número de modelos es elevado y se requiere una alta reducción de dimensionalidad [93].



Hasta ahora, los procesos para la fusión de datos se centran principalmente en los niveles medio y bajo [35]. Por lo general, el criterio principal para una óptima fusión de datos es el poder predictivo de las distintas plataformas de forma individual. Sin embargo, incluso las plataformas con bajo poder predictivo individual pueden aumentar en gran medida la precisión tras la fusión porque se considera la relación multivariada entre plataformas concatenadas. La aplicación de la fusión de datos es muy beneficiosa y es un tema aplicado en diferentes campos, como en el análisis de procesos industriales, el control integral de la calidad y el análisis de sensores y comunicaciones [93].

En los últimos años, la fusión de datos en estudios metabolómicos ha mostrado un gran interés, reflejándose en el aumento de publicaciones en distintas matrices alimentarias como vino [96], cerveza [97] o aceite [98]. La mayoría de los estudios basados en fusión de datos provienen de técnicas espectroscópicas y espectrométricas: infrarrojo cercano (NIR), infrarrojo medio (MIR), ultravioleta-visible (UV-Vis), Raman, NMR o MS [74]. Hasta la fecha de la realización de esta Tesis Doctoral no se han publicado estudios de fusión de datos de GC-MS, LC-MS y NMR en matrices alimentarias. Se considera que dichos estudios basados en información espectral y complementaria como la proporcionada por las tres técnicas mencionadas constituirán una importante plataforma soporte de estudios metabolómicos en sistemas complejos.

1.3 Resonancia magnética nuclear cuantitativa

1.3.1 Introducción a la NMR cuantitativa

Durante años, la principal aplicación de la NMR ha estado ligada a estudios de elucidación estructural. Todos aquellos isótopos magnéticamente activos (como hidrógeno-1, carbono-13, nitrógeno-15, flúor-19, fósforo-31, etc.) permiten el estudio estructural de gran variedad de sustancias gracias a la radiación electromagnética generada por los mismos bajo la influencia de un campo magnético externo (B_0) (Figura 2). Además, el amplio catálogo de disolventes deuterados disponibles hoy en día, indispensables para la calibración del campo magnético (*shimming*) usando la frecuencia de resonancia de los átomos de deuterio presentes en la muestra, permite la resolución de un alto número de problemas analíticos.

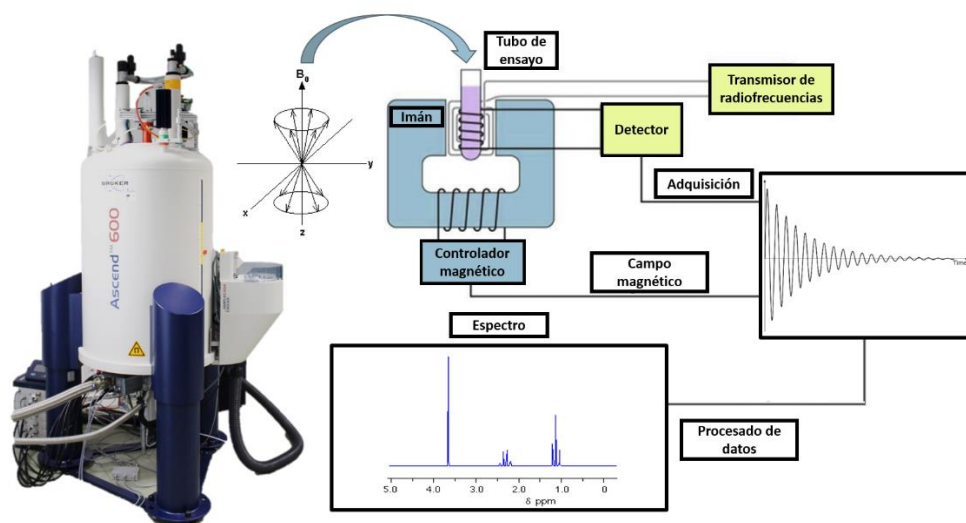


Figura 2. Esquema general de un equipo de resonancia magnética nuclear moderno



Las señales producidas por los pulsos magnéticos aplicados (P_1) seguida de los tiempos de relajación (D_1) en cada uno de los *scans* adquiridos (NS) aportan gran cantidad de información química: desplazamientos químicos (δ) en partes por millón (ppm), procesos de relajación, acoplamientos escalares, acoplamientos dipolares, intercambios químicos, procesos dinámicos o difusión molecular, entre otros [99,100].

Una vez obtenida la señal es necesario realizar una serie de operaciones básicas: el pre-procesado de los datos. Éstas incluyen la corrección de la línea base, calibración del desplazamiento químico, deconvolución, integración de las señales, etc. La selección de los parámetros de adquisición, tratamientos de pre-procesado de datos y la selección de la matriz de estudio son elementos cruciales para el éxito de un análisis cuantitativo (Figura 3) [49].

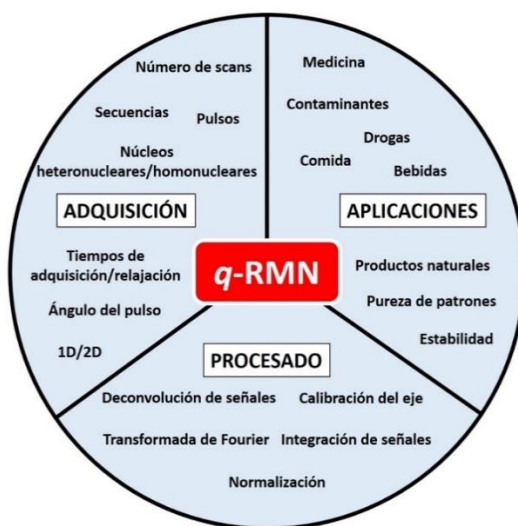


Figura 3. Esquema de los distintos parámetros de adquisición, tratamientos de pre-procesado y aplicaciones posibles que requiere un análisis de qNMR

En los últimos años la q -NMR ha ganado terreno como técnica analítica tanto en ámbito técnico como de investigación [101,102]. Bajo ciertas condiciones, la proporcionalidad que existe entre el valor absoluto de la integral de una señal obtenida con el número de núcleos presentes, da lugar a análisis cuantitativos en matrices complejas con excelentes resultados. Estas propiedades han provocado un incremento exponencial de aplicaciones de la q NMR en muy diversas matrices como extractos orgánicos [103], drogas [104,105], fluidos corporales [106], productos naturales [32], suplementos dietéticos [33] o fármacos [34].

Desde la primera publicación del uso de la q NMR en la década de los años 60 [107], la mayoría de las aplicaciones de la NMR han estado basadas en la observación de núcleos con espín $I=1/2$. Si bien es cierto que hasta las tres últimas décadas la q -NMR no ha suscitado un alto interés científico debido a las limitaciones instrumentales que presentaba. Hoy en día es una de las técnicas que mayor crecimiento ha experimentado en cuanto a número de aplicaciones en campos como la certificación de pureza de patrones, el control de los productos naturales y la investigación farmacéutica [108]. Dicho crecimiento está asociado, además de a las mejoras instrumentales (p. ej. nuevos materiales superconductores), a la aplicación de métodos multipulsos, hiperpolarización o el desarrollo de los métodos bidimensionales, dando lugar a un *boom* tecnológico en multitud de ámbitos en los últimos años [109]. Asimismo, la q -NMR es considerada como una técnica primaria de medida, al permitir el análisis directo de las relaciones entre el área de la señal de un analito y la de un estándar de concentración conocida [110].



Basándonos en datos extraídos de la web global de publicaciones científicas *Scopus* [111], desde el año 2000, dicha técnica ha mostrado un incremento de más de un 200% respecto al número de publicaciones incluyendo las palabras clave “*Quantitative NMR*”, “*q-NMR*” y “*qNMR*” (Figura 4).

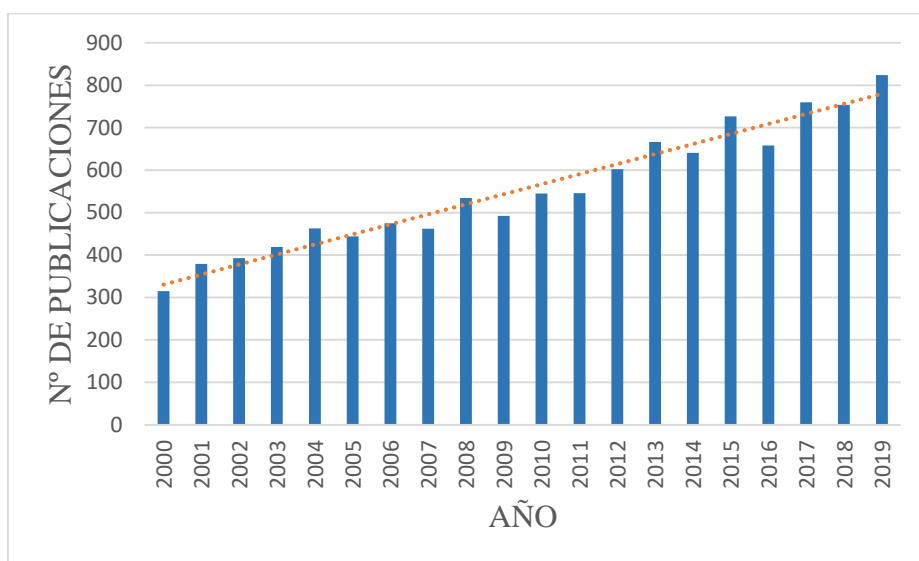


Figura 4. Resultados de búsqueda en Scopus del número de publicaciones por año para las palabras clave “*quantitative NMR*”, “*q-NMR*” y “*qNMR*”

La mayoría de las aplicaciones cuantitativas de la *qNMR* publicadas están enfocadas al estudio de núcleos dipolares, espín $I = 1/2$, como ^1H , ^3H , ^{13}C , ^{15}N , ^{19}F , ^{29}Si , ^{31}P , ^{57}Fe y algunos otros. Uno de los problemas asociados a la mayoría de heteronúcleos, aquellos diferentes al ^1H , es su baja sensibilidad. Sin embargo, algunos como el ^{19}F o ^{31}P presentan abundancias relativas altas (núcleos sensibles), además de largos radios

giromagnéticos (γ) y momentos magnéticos (μ) elevados [112]. Estas características ofrecen así una idónea sensibilidad en las adquisiciones de los núcleos. En ocasiones donde no se alcanza una sensibilidad mínima necesaria, dicho problema puede abordarse a través de una detección indirecta, registrando la intensidad de protones acoplados al analito objetivo.

Otros inconvenientes añadidos que pueden encontrarse en núcleos dipolares son sus elevados anchos de ventana, como el caso del ^{195}Pt (8000 ppm) [113], con su consecuente pérdida de resolución, o con tiempos de relajación extremadamente largos (varios minutos), como sucede con el ^{57}Fe [114]. La adición de compuestos paramagnéticos, como el $\text{Cr}(\text{acetilacetato})_3$, puede reducir el tiempo de relajación hasta valores aceptables, con el inconveniente de posibles alteraciones espectrales [115].

El estudio de los compuestos heteronucleares, desde sus primeras aplicaciones cuantitativas en los años 70, ha experimentado un incremento porcentual notable durante los años transcurridos en el presente siglo, si bien su número es aún no es muy elevado, poniendo de manifiesto el potencial de crecimiento e interés en la técnica (Figura 5).

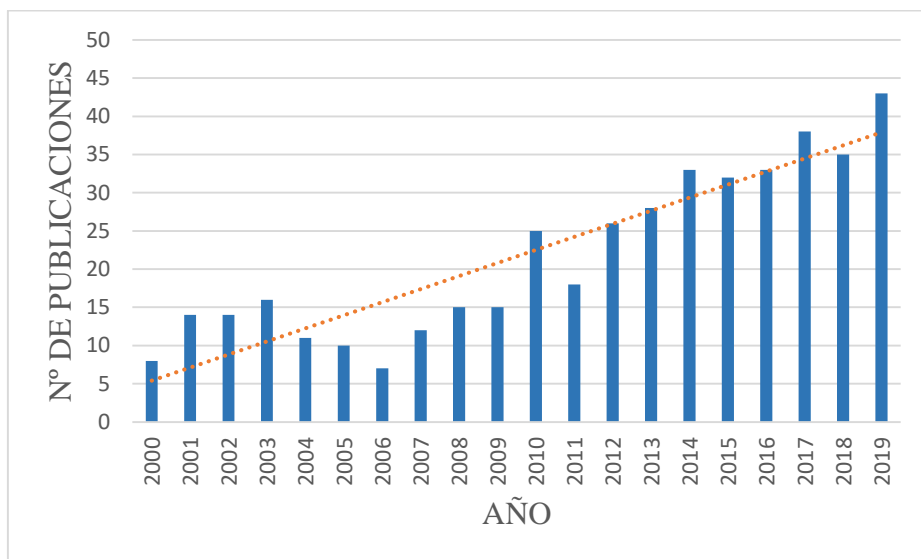


Figura 5. Resultados de búsqueda en Scopus del número de publicaciones por año para las palabras clave "Heteronuclear quantitative NMR", " Heteronuclear q-NMR" y " Heteronuclear qNMR"

Concretamente, uno de los casos que más potencial está mostrando es el estudio de compuestos con núcleos de fósforo-31. Este elemento tiene un importante rol en procesos fisiológicos, formando parte incluso del DNA en forma de ésteres ácido fosfóricos [116]. Además, este tipo de compuestos se encuentra muy presente en el medio ambiente, normalmente en compuestos con fósforo enlazado a heteronúcleos como azufre, oxígeno o nitrógeno, siendo ampliamente utilizados como plaguicidas, lubricantes, materiales plásticos o drogas, entre otros [117]. Por ello, el estudio y seguimiento de éste tipo de compuestos mediante técnicas como la ^{31}P -qNMR es crucial para garantizar el control de los mismos.

CAPÍTULO PRIMERO. MARCO TEÓRICO

Por otro lado, el estudio de las aplicaciones cuantitativas basadas en núcleos cuadrupolares ($I > 1/2$) apenas se ha investigado, pese a que la mayoría de los compuestos heteronucleares pertenecen a este grupo. Además, muchos de estos compuestos son metales, metales de transición o elementos con importantes funciones biológicas, que en la mayoría de los casos están enlazados con iones en soluciones acuosas.

El principal motivo de haber sido poco estudiados son las grandes diferencias fundamentales en los factores gobernantes del proceso de relajación magnética nuclear entre los núcleos cuadrupolares y dipolares: poseen momentos eléctricos cuadrupolares (Q) que normalmente dan lugar a un considerable ensanchamiento de señales. En algunos casos, los desplazamientos químicos entre núcleos no equivalentes cuadrupolares pueden ser muy grandes (p. ej. 20000 ppm para el ^{59}Co) [118].

Algunos núcleos cuadrupolares, como el boro-11 (alta abundancia del 80.4% y un espín= 3/2), presentan una óptima sensibilidad y resolución en un espectro asequible de 200 ppm de ancho. Las posibles interferencias producidas en este tipo de análisis, por las estructuras cristalinas borosilicatadas de los tubos de ensayo tienen que evitarse (p. ej. con el uso de tubos de medida de cuarzo) [119].

El uso de compuestos borados tiene un reconocido interés científico como análogos de moléculas orgánicas, adquiriendo propiedades similares. Ello permite su aplicación en campos como la medicina, siendo usados en tratamientos contra el cáncer o la agricultura en forma de gran variedad de formulados de biocidas, insecticidas, herbicidas o fungicidas [120,121]. Por ello, la aplicación de la $q\text{NMR}$ en compuestos borados se presenta como una alternativa con gran potencial analítico.



Ya desde el primer empleo analítico de la q NMR basada en núcleos cuadrupolares por *Maki et al* [122] en el año 2017, esta técnica ha demostrado su gran capacidad de aplicación a este tipo de núcleos, abriendo las puertas al estudio analítico de los mismos. Los resultados obtenidos son esperanzadores y dan lugar a nuevas investigaciones de aplicación de la q NMR a nuevos analitos y matrices.

Si bien es cierto que hoy en día aún existen importantes desafíos que la q NMR debe de superar, entre ellos su baja sensibilidad respecto a otras técnicas. Para dar solución a ello, en los últimos años se han conseguido grandes avances en el desarrollo de métodos de adquisición más rápidos, uso de sondas criogénicas enfriadas con nitrógeno, acoplamientos con LC o el uso simultáneo de varios receptores [123]. Dichas mejoras convierten a la q NMR en una de las principales técnicas analíticas aplicables a todo tipo de sustancias activas [109]. Las ventajas y desventajas de las técnicas previamente descritas se encuentran resumidas en la Tabla 8.

Tabla 8. Ventajas y desventajas de las principales técnicas de análisis cuantitativo

Técnica	Ventajas	Desventajas
Espectroscopía de absorción atómica (EAA)	<ul style="list-style-type: none"> ✓ Buenos límites de cuantificación ✓ Sencillo manejo ✓ Bajo coste económico 	<ul style="list-style-type: none"> ✗ Requiere patrones para la cuantificación mediante rectas de calibrado ✗ Técnica destructiva ✗ Análisis poco selectivo (determinación elemental) ✗ Análisis elemental uno a uno

CAPÍTULO PRIMERO. MARCO TEÓRICO

Técnica	Ventajas	Desventajas
GC-MS	<ul style="list-style-type: none"> ✓ Tiempo de retención reproducible ✓ Excelentes límites de cuantificación ✓ Buena precisión y resolución 	<ul style="list-style-type: none"> ✗ Requiere extracción ✗ Requiere patrones para la cuantificación mediante rectas de calibrado ✗ Requiere separación cromatográfica ✗ Puede requerir derivatización ✗ Técnica destructiva ✗ Requiere volatilidad de los analitos ✗ Alto coste económico
IC	<ul style="list-style-type: none"> ✓ Buenos límites de cuantificación ✓ Buena precisión y resolución 	<ul style="list-style-type: none"> ✗ Requiere patrones para la cuantificación mediante rectas de calibrado ✗ Requiere separación cromatográfica ✗ Análisis poco selectivo (determinación de iones y azúcares)
ICP-MS	<ul style="list-style-type: none"> ✓ Excelentes límites de cuantificación ✓ Buena precisión y resolución ✓ Tiempos de análisis cortos 	<ul style="list-style-type: none"> ✗ Alto coste económico ✗ Técnica destructiva ✗ Requiere patrones para la cuantificación mediante rectas de calibrado





Técnica	Ventajas	Desventajas
LC-MS	<ul style="list-style-type: none"> ✓ Tiempo de retención reproducible ✓ Excelentes límites de cuantificación ✓ Buena precisión y resolución 	<ul style="list-style-type: none"> ✗ Requiere extracción ✗ Requiere patrones para la cuantificación mediante rectas de calibrado ✗ Requiere separación cromatográfica ✗ Puede requerir derivatización ✗ Técnica destructiva ✗ Requiere cierta polaridad de los analitos objetivo ✗ Alto coste económico
NMR	<ul style="list-style-type: none"> ✓ Rapidez ✓ Alta repetitividad ✓ Alta reproducibilidad ✓ No destructiva ✓ Mínima preparación de muestra ✓ Robusta ✓ No requiere patrón específico ✓ Varias determinaciones simultáneas ✓ Aplicable a muestras complejas 	<ul style="list-style-type: none"> ✗ Baja sensibilidad ✗ Alto coste económico ✗ Rigurosa interpretación de los espectros ✗ Requiere disolventes deuterados

1.3.2 Proceso analítico qNMR: pretratamiento y adquisición

La selección de un disolvente deuterado es el primer parámetro a decidir en la preparación de las muestras dado que se requiere la calibración para la estabilidad del campo magnético en los instrumentos de NMR modernos. Dicha selección está basada en diferentes criterios, siendo el primero que el analito sea soluble en el mismo a la concentración requerida para el estudio. Esto dependerá de diferentes factores como la sensibilidad del núcleo o la temperatura del experimento. En el caso de cuantificación interna, es imprescindible que además el estándar interno (IS) seleccionado sea también soluble en dicho disolvente.

Ante la posible presencia de partículas sólidas que puedan distorsionar el campo magnético, es necesario filtrar los extractos de muestras o disoluciones de manera previa al análisis. Una buena práctica es el uso de filtros, resinas quelantes o lana sintética en el interior de una pipeta Pasteur, previamente lavada con el disolvente deuterado. En aquellos análisis donde se requiere de una resolución máxima de las señales, las muestras pueden ser desgasificadas con la finalidad de eliminar el posible aire presente en las disoluciones. Otro factor a tener en cuenta ha de ser el desplazamiento químico que tendrán las señales producidas por el disolvente, anticipándose y tratando de evitar posibles solapamientos con los analitos objeto de estudio.

Respecto a la cuantificación en NMR puede realizarse mediante dos métodos: cuantificación externa, es decir mediante uso de rectas de calibrado con los patrones del analito objetivo, y cuantificación interna, basada en el uso de patrones del núcleo magnéticamente observable, con el cuál se va a trabajar, con la ventaja de que las variaciones instrumentales y experimentales van a ser mínimas. Ésta última es la



modalidad más comúnmente usada hoy en día. El uso de IS con desplazamientos químicos a 0 ppm (p. ej. tetrametilsilano- d_4 , TMS, o 3-[trimetilsilil]-propionato- d_4 de sodio, TMSP) es una de las prácticas más extendidas [124].

La consideración de un patrón analítico como apto para una cuantificación mediante q NMR precisa de unos requisitos básicos. A tal fin, un IS debe de cumplir con las siguientes condiciones [125]:

- Ser fácilmente accesible comercialmente con un alto grado de pureza
- No ser un analito higroscópico ni volátil en condiciones de temperatura ambiente
- Ser poco reactivo químicamente y toxicológicamente
- Producir espectros magnéticos sencillos y con pocas señales para evitar un solapamiento magnético de las mismas
- Presentar desplazamientos químicos que cubran distintas regiones del espectro, permitiendo varias opciones de cuantificación
- Presentar una alta solubilidad en algún disolvente deuterado.
- Tener un precio razonable
- Contar con tiempos de relajación cortos (del orden de $D_1= 5$ a 7 veces el tiempo de relajación longitudinal, T_1)

Hoy en día existen gran número de patrones destinados a la q NMR disponibles comercialmente y ampliamente “aceptados” como patrones de referencia en la comunidad científica (p. ej. acetanilida, dimetil sulfona, 3,4,5-tricloropiridina, etc)[123]. Para casos donde se requiere la

cuantificación de núcleos menos habituales, pueden usarse listas de frecuencias de referencia para núcleos específicos. Una de las más aplicadas es la lista de frecuencias de materiales de referencia para todos los núcleos relativa a la señal $^1\text{H-NMR}$ del TMS, desarrollada y publicada por la Unión Internacional de Química Pura y Aplicada (IUPAC) [126,127].

A la hora de la cuantificación a través de una calibración interna, el cálculo de la concentración debe de realizarse teniendo en cuenta los pesos moleculares del analito objetivo X (M_x) y del IS (M_y), así como el peso medido con exactitud del IS (m_y) (Ecuación 8). La obtención de buenos resultados a la hora de la cuantificación se verá determinada por intensidades semejantes del IS respecto a la del analito objetivo (p. ej, evitando distorsiones del espectro por sobresaturación del IS). Con los valores de las integrales relativas a X (i_x), de Y (i_y) y el cálculo de la relación molar n_x/n_y , es posible la determinación de la cantidad de analito (m_x) por la siguiente relación [128]:

$$m_x = \frac{i_x/n_x}{i_y/n_y} * \frac{M_x}{M_y} * m_y$$

Ecuación 8. Ecuación para el cálculo de la concentración del analito objetivo en valores absolutos mediante qNMR y calibración interna.

El volumen de muestra requerido para los análisis mediante qNMR debe de ser suficiente para cubrir toda la longitud del rango de detección del equipo, evitando así las distorsiones locales que el campo magnético puede sufrir en presencia de aire. Para un tubo estándar de 5 mm de diámetro, un volumen de 500-600 μL es suficiente. Es necesario el uso de tapones para conseguir un aislamiento óptimo de la muestra, siendo el tetrafluoretileno el material más extendido [129].





Una vez la muestra está preparada para ser analizada, el equipo de NMR debe de ser puesto a punto. El desarrollo de nuevos métodos para el análisis cuantitativo por NMR requiere de unas consideraciones previas. Así es necesario ajustar distintos parámetros, como el P_1 , D_1 y la ganancia del receptor (RG) a fin de conseguir una correcta cuantificación y reproducibilidad. El uso de pulsos magnéticos de 90° y un tiempo de repetición de pulso igual a $5 \times T_1$, es una de las prácticas más comunes para asegurar la relajación completa de los núcleos excitados, aunque se puede lograr una cuantificación confiable cuando se utilizan ángulos de pulso más pequeños y tiempos de repetición más cortos. Se requiere una precaución especial para evitar problemas de rango dinámico causados por grandes diferencias de concentración entre varios analitos [58].

En algunos casos puede ser útil realizar un desacoplamiento de carbono para evitar las interferencias producidas durante la adquisición debida a la presencia de satélites ^{13}C . Para ello pueden usarse esquemas de pulsos para desacoplamiento restringido inverso de carbono, evitando así sobrecalentamiento de las muestras y suprimiendo el efecto nuclear Overhauser (*nuclear Overhauser effect*, NOE), que distorsiona la intensidad de las señales mediante magnetización entre núcleos próximos en el espacio.

Por último, tal como indican las guías internacionales vigentes [130], el estado del equipo de NMR debe de verificarse sistemáticamente chequeando la resolución, el ancho de las señales, la asimetría de las mismas y la señal relación entre la señal y el ruido (S/N). En la Tabla 9 se encuentran recogidos los principales parámetros necesarios a tener en cuenta en la aplicación de la $q\text{NMR}$.

Tabla 9. Principales consideraciones a tener en cuenta en determinaciones mediante q NMR

Planteamiento	Pretratamiento	Adquisición	Procesado de datos
<ul style="list-style-type: none"> • Modo de cuantificación (interna o externa) • Núcleo objeto de estudio • Selección del disolvente deuterado • Selección de patrón analítico 	<ul style="list-style-type: none"> • Disolución completa de IS • Disolución completa de matriz objetivo • Ausencia de partículas o burbujas • Evitar posible saturación del RG (p. ej. mediante diluciones) 	<ul style="list-style-type: none"> • Secuencia de pulsos • Ángulo del pulso (generalmente 90° o menos) • Calibración de la línea base • RG* • NS** • D_1*** 	<ul style="list-style-type: none"> • Transformada de Fourier • Corrección de la línea base • Calibración del espectro (p. ej. con TMS a 0.0 ppm) • S/N • Correcta integración de las señales (evitando aquellas con solapamiento)

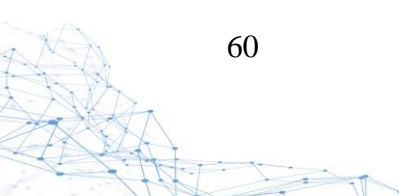
*Ganancia del Receptor (*Receiver Gain*)

**Número de *scans*

***Tiempo de relajación

1.3.3 Validación del método

La demanda de validaciones analíticas que garanticen una correcta evaluación de los métodos desarrollados es uno de los retos a los que se enfrenta la q -NMR. Hoy en día, la falta de métodos oficiales de q -NMR supone una seria limitación para la explotación del potencial de la NMR tanto en análisis cuantitativos de un solo componente como en determinaciones múltiples [131].





Actualmente, algunas directrices vigentes, como la Guía SANTE o la Decisión (EC) 2002/657, son muy destacadas y ampliamente empleadas para la validación de métodos analíticos. Estos documentos están mayormente orientados a la determinación de plaguicidas en matrices alimentarias mediante técnicas cromatográficas acopladas a analizadores de MS [132,133].

No obstante, existen guías con directrices de carácter general sobre la validación de métodos que pueden ser aplicadas a la q -NMR, como la Guía de laboratorio para la validación de métodos y temas relacionados, elaborada por la red de organizaciones europeas relativas a la química (Eurachem) [134]. Dicha guía define la validación de un método como el proceso para definir un requisito analítico, confirmándose a tal modo que cuenta las capacidades necesarias para las aplicaciones requeridas.

Otra guía, la de la Administración de Medicamentos y Alimentos (*Food and Drug Administration*, FDA), considera diversos parámetros de validación para garantizar que los métodos analíticos sean adecuados, como exactitud, rango de trabajo, curva de calibración, muestras de control de calidad, selectividad, exactitud y estabilidad del analito objetivo en la matriz. El enfoque de dichas recomendaciones está orientado a la determinación cuantitativa de medicamentos y sus metabolitos, proteínas o marcadores en muestras de origen biológico. En cuanto a criterios de confirmación, la FDA se centra en el uso de técnicas espectrométricas como herramienta universal [135].

Con la finalidad de unificar todos los criterios de validación y adaptarlos a su aplicación en el desarrollo de métodos mediante NMR, Eurolab ha redactado una Guía para el desarrollo y validación de métodos con

carácter cualitativo y cuantitativo [136]. En ellas, sugieren una serie de requisitos mínimos a evaluar, con la finalidad principal del control y seguimiento de alimentos, incluyendo criterios de confirmación específicos de q NMR.

En términos generales, y dada la alta capacidad elucidativa de la NMR, una verificación estructural de un analito concreto puede darse mediante la asignación de todas las señales obtenidas. En casos concretos donde la confirmación directa no es posible (p. ej. alta complejidad del espectro), es recomendable el uso de análisis de correlación bidimensionales (p. ej. H, H-COSY, HSQ, HMBC, etc.). Existen estrategias adicionales que pueden ser aplicables, como la adquisición de diferentes núcleos activos, el uso de *software* de predicción (p. ej. *PERCH* [137], *Structure Elucidator* [138], *CCASA* [139], *StrucEluc* [139], etc.) o uso de librerías espectrales (p.ej. *HMDB* [140], *MetaboLights* [141], *MetaboAnalyst* [142], etc.). En el caso de muestras complejas con solapamiento de señales pueden ser útiles las adquisiciones mediante experimentos de difusión molecular, permitiendo el estudio individual de los mismos [143].

Como se puede observar, la alta capacidad de identificación de la NMR permite confirmaciones mucho más simples que las técnicas espectrométricas, así como reducción de tiempo de análisis y costes al poder usarse los mismos análisis tanto para identificación como cuantificación.

Los parámetros de validación indicados en la guía Eurolab para métodos de q NMR son los siguientes:

- Veracidad: Estimada mediante comparación de las purezas de varios materiales de referencia certificados o mediante



comparación de valores de recuperación (%) de muestras de concentraciones conocidas.

- **Precisión:** Estimada mediante la determinación de múltiples mediciones en diferentes intervalos de tiempo de varias réplicas independientes de una misma muestra.
- **Selectividad:** En la técnica de NMR, se puede garantizar la selectividad del método cuando no se da solapamiento de las señales del analito objetivo en la matriz objeto de estudio.
- **Límite de detección (LOD):** es la concentración más baja a la cual el analito puede ser detectado (señal con valor de $S/N \geq 3$).
- **Límite de cuantificación (LOQ):** es la concentración más baja a la cual el analito puede cuantificarse con una precisión adecuada.
- **Linealidad:** Debe verificarse mediante el uso de al menos 5 concentraciones diferentes de la sustancia objetivo.
- **Rango de trabajo:** El rango de trabajo de un método de q NMR generalmente comienza desde el LOQ hasta la concentración de la máxima solubilidad del analito (límite superior) en una matriz dada.
- **Robustez:** Deben considerarse y examinarse las posibles influencias de diferentes factores (por ejemplo, valores de pH, altas concentraciones de sal de la matriz de la muestra, productos químicos reactivos).
- **Incertidumbre (U):** Error sistemático de múltiples determinaciones de concentración de analito en la muestra. Cuando no se detecta un error sistemático, la incertidumbre de medición es igual a la precisión. En otros casos, la concentración

de analito calculada debe ajustarse mediante valores de recuperación estimados.

Los parámetros definidos por Eurolab pueden complementarse con las especificaciones dadas por las guías previamente descritas [133]. De ésta forma, la determinación del LOQ puede verse definida como la concentración más baja del analito objetivo que genera valores de veracidad y precisión adecuados, así como la precisión puede estimarse en términos de desviación estándar relativa (RSD) entre análisis intradía e interdía. [135].

Una práctica muy extendida para la determinación de la incertidumbre del método es aplicando el denominado enfoque *bottom-up* (desde abajo hacia arriba). Para ello, la idoneidad del método puede ser calculada mediante el cálculo del error sistemático a la concentración más baja y más alta del rango de trabajo [134].



1.4 Bibliografía

- [1] F. Capozzi, A. Bordoni, Foodomics: A new comprehensive approach to food and nutrition, *Genes Nutr.* 8 (2013) 1–4. doi:10.1007/s12263-012-0310-x.
- [2] C. Ibáñez, C. Simó, V. García-Cañas, A. Cifuentes, M. Castro-Puyana, Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics: A review, *Anal. Chim. Acta.* 802 (2013) 1–13. doi:10.1016/j.aca.2013.07.042.
- [3] Y. Lu, X.M. Zhao, Z. Hu, L. Wang, F. Li, lc–ms-based metabolomics in the study of drug-induced liver injury, *Curr. Pharmacol. Reports.* 5 (2019) 56–67. doi:10.1007/s40495-018-0144-3.
- [4] H.K. Biesalski, L.O. Dragsted, I. Elmadfa, R. Grossklaus, M. Müller, D. Schrenk, P. Walter, P. Weber, Bioactive compounds: Definition and assessment of activity, *Nutrition.* 25 (2009) 1202–1205. doi:10.1016/j.nut.2009.04.023.
- [5] J. Nielsen, S. Oliver, The next wave in metabolome analysis, *Trends Biotechnol.* 23 (2005) 544–546. doi:10.1016/j.tibtech.2005.08.005.
- [6] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, D.B. Kell, Metabolomics by numbers: Acquiring and understanding global metabolite data, *Trends Biotechnol.* 22 (2004) 245–252. doi:10.1016/j.tibtech.2004.03.007.

- [7] G.J. Patti, O. Yanes, G. Siuzdak, Innovation: Metabolomics: the apogee of the omics trilogy, *Nat. Rev. Mol. Cell Biol.* 13 (2012) 263–269. doi:10.1038/nrm3314.
- [8] L.M. McShane, M.M. Cavenagh, T.G. Lively, D.A. Eberhard, W.L. Bigbee, P.M. Williams, J.P. Mesirov, M.Y.C. Polley, K.Y. Kim, J. V. Tricoli, J.M.G. Taylor, D.J. Shuman, R.M. Simon, J.H. Doroshow, B.A. Conley, Criteria for the use of omics-based predictors in clinical trials, *Nature.* 502 (2013) 317–320. doi:10.1038/nature12564.
- [9] O. Fiehn, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, *Comp. Funct. Genomics.* 2 (2001) 155–168. doi:10.1002/cfg.82.
- [10] R.A. Carter, K. Pan, E.W. Harville, S. McRitchie, S. Sumner, Metabolomics to reveal biomarkers and pathways of preterm birth: a systematic review and epidemiologic perspective, *Metabolomics.* 15 (2019) 1–27. doi:10.1007/s11306-019-1587-1.
- [11] R. Chaleckis, I. Meister, P. Zhang, C.E. Wheelock, Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics, *Curr. Opin. Biotechnol.* 55 (2019) 44–50. doi:10.1016/j.copbio.2018.07.010.
- [12] M.E. Dasenaki, N.S. Thomaidis, Quality and authenticity control of fruit juices-a review, 2019. doi:10.3390/molecules24061014.
- [13] J. Riedl, S. Esslinger, C. Fahl-Hassek, Review of validation and reporting of non-targeted fingerprinting approaches for food authentication, *Anal. Chim. Acta.* 885 (2015) 17–32. doi:10.1016/j.aca.2015.06.003.



- [14] R. Consonni, D. Polla, L.R. Cagliani, Organic and conventional coffee differentiation by NMR spectroscopy, *Food Control*. 94 (2018) 284–288. doi:10.1016/J.FOODCONT.2018.07.013.
- [15] K. Navratilova, V. Hrbek, F. Kratky, K. Hurkova, M. Tomaniova, J. Pulkrabova, J. Hajslova, Green tea: Authentication of geographic origin based on UHPLC-HRMS fingerprints, *J. Food Compos. Anal.* 78 (2019) 121–128. doi:10.1016/j.jfca.2019.02.004.
- [16] S. Ghisoni, L. Lucini, F. Angilletta, G. Rocchetti, D. Farinelli, S. Tombesi, M. Trevisan, Discrimination of extra-virgin-olive oils from different cultivars and geographical origins by untargeted metabolomics, *Food Res. Int.* 121 (2019) 746–753. doi:10.1016/j.foodres.2018.12.052.
- [17] A. Da Ros, D. Masuero, S. Riccadonna, K. Brkić Bubola, N. Mulinacci, F. Mattivi, I. Lukić, U. Vrhovsek, complementary untargeted and targeted metabolomics for differentiation of extra virgin olive oils of different origin of purchase based on volatile and phenolic composition and sensory quality, *Molecules*. 24 (2019) 2896. doi:10.3390/molecules24162896.
- [18] G. Danezis, C. Theodorou, T. Massouras, E. Zoidis, I. Hadjigeorgiou, C.A. Georgiou, Greek graviera cheese assessment through elemental metabolomics—Implications for authentication, safety and nutrition, *Molecules*. 24 (2019). doi:10.3390/molecules24040670.
- [19] A. Ali, M. Khan, R. Sharif, M. Mujtaba, S.-J. Gao, Sugarcane

- Omics: An Update on the Current Status of Research and Crop Improvement, *Plants*. 8 (2019) 344. doi:10.3390/plants8090344.
- [20] R. Consonni, F. Bernareggi, L.R. Cagliani, NMR-based metabolomic approach to differentiate organic and conventional Italian honey, *Food Control*. 98 (2019) 133–140. doi:10.1016/j.foodcont.2018.11.007.
- [21] Q.Q. Ma, X.D. Wang, Y.Y. Cui, N.N. Zhang, J.G. Qin, Z.Y. Du, L.Q. Chen, Untargeted GC-MS metabolomics reveals metabolic differences in the Chinese mitten-hand crab (*Eriocheir sinensis*) fed with dietary palm oil or olive oil, *Aquac. Nutr.* 24 (2018) 1623–1637. doi:10.1111/anu.12798.
- [22] L.A. da Silva, D.L. Flumignan, A.G. Tininis, H.R. Pezza, L. Pezza, Discrimination of Brazilian lager beer by ¹H NMR spectroscopy combined with chemometrics, *Food Chem.* 272 (2019) 488–493. doi:10.1016/j.foodchem.2018.08.077.
- [23] T. Kuballa, T.S. Brunner, T. Thongpanchang, S.G. Walch, D.W. Lachenmeier, Application of NMR for authentication of honey, beer and spices, *Curr. Opin. Food Sci.* 19 (2018) 57–62. doi:10.1016/j.cofs.2018.01.007.
- [24] S. Canas, phenolic composition and related properties of aged wine spirits: influence of barrel characteristics. A Review, *Beverages*. 3 (2017) 55. doi:10.3390/beverages3040055.
- [25] L. Gougeon, G. da Costa, I. Le Mao, W. Ma, P.L. Teissedre, F. Guyon, T. Richard, wine analysis and authenticity using 1h-nmr metabolomics data: application to chinese wines, *Food Anal. Methods*. (2018) 1–10. doi:10.1007/s12161-018-1310-2.



- [26] M. Stupak, V. Kocourek, I. Kolouchova, J. Hajslova, Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC–MS, *Food Control*. 80 (2017) 307–313. doi:10.1016/j.foodcont.2017.05.008.
- [27] M.P. Callao, I. Ruisánchez, An overview of multivariate qualitative methods for food fraud detection, *Food Control*. 86 (2018) 283–293. doi:10.1016/j.foodcont.2017.11.034.
- [28] M. Efenberger-Szmechtyk, A. Nowak, D. Kregiel, Implementation of chemometrics in quality evaluation of food and beverages, *Crit. Rev. Food Sci. Nutr.* 58 (2018) 1747–1766. doi:10.1080/10408398.2016.1276883.
- [29] E. Hong, S.Y. Lee, J.Y. Jeong, J.M. Park, B.H. Kim, K. Kwon, H.S. Chun, Modern analytical methods for the detection of food fraud and adulteration by food category, *J. Sci. Food Agric.* 97 (2017) 3877–3896. doi:10.1002/jsfa.8364.
- [30] R. Romero-González, A. Garrido Frenich, application of hrms in pesticide residue analysis in food from animal origin, *Appl. High Resolut. Mass Spectrom. Food Saf. Pestic. Residue Anal.* (2017) 203–232. doi:10.1016/B978-0-12-809464-8.00007-5.
- [31] H. Awad, M.M. Khamis, A. El-Aneed, Mass spectrometry, review of the basics: Ionization, *Appl. Spectrosc. Rev.* 50 (2015) 158–175. doi:10.1080/05704928.2014.954046.
- [32] C. Simmler, J.G. Napolitano, J.B. McAlpine, S.N. Chen, G.F. Pauli, Universal quantitative NMR analysis of complex natural

- samples, *Curr. Opin. Biotechnol.* 25 (2014) 51–59. doi:10.1016/j.copbio.2013.08.004.
- [33] P.M. Le, C. Milande, E. Martineau, P. Giraudeau, J. Farjon, Quantification of natural products in herbal supplements: A combined NMR approach applied on goldenseal, *J. Pharm. Biomed. Anal.* 165 (2019) 155–161. doi:10.1016/j.jpba.2018.11.062.
- [34] A. Owczarek, A. Kłys, M.A. Olszewska, A validated ¹H qNMR method for direct and simultaneous quantification of esculin, fraxin and (–)-epicatechin in Hippocastani cortex, *Talanta*. 192 (2019) 263–269. doi:10.1016/j.talanta.2018.09.036.
- [35] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, Y. Liang, Chemometric methods in data processing of mass spectrometry-based metabolomics: A review, *Anal. Chim. Acta*. 914 (2016) 17–34. doi:10.1016/j.aca.2016.02.001.
- [36] K.H. Liland, Multivariate methods in metabolomics - from pre-processing to dimension reduction and statistical analysis, *TrAC - Trends Anal. Chem.* 30 (2011) 827–841. doi:10.1016/j.trac.2011.02.007.
- [37] 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 Food Sci. Technol.* 20 (2009) 557–566. doi:10.1016/j.tifs.2009.07.002.
- [38] R. Vázquez-Fresno, S. Tulipani, O. Khymenets, M. Urpi-Sarda, M. Garcia-Aloy, M. Rabassa, M. Boto-Ordoñez, M. Rotches-Ribalta, R. Llorach, C. Andres-Lacueva, emerging applications of



- metabolomics to polyphenols and cvd biomarker discovery, *Polyphenols Hum. Heal. Dis.* 2 (2013) 1025–1044. doi:10.1016/B978-0-12-398456-2.00079-7.
- [39] B. Worley, R. Powers, multivariate analysis in metabolomics, *Curr. Metabolomics.* 1 (2012) 92–107. doi:10.2174/2213235X130108.
- [40] K. Lepenioti, A. Bousdekis, D. Apostolou, G. Mentzas, Prescriptive analytics: Literature review and research challenges, *Int. J. Inf. Manage.* 50 (2020) 57–70. doi:10.1016/j.ijinfomgt.2019.04.003.
- [41] B.A. Boughton, B. Hamilton, spatial metabolite profiling by matrix-assisted laser desorption ionization mass spectrometry Imaging, in: *Adv. Exp. Med. Biol.*, 2017: pp. 291–321. doi:10.1007/978-3-319-47656-8_12.
- [42] B. Worley, R. Powers, multivariate analysis in metabolomics, *Curr. Metabolomics.* 1 (2012) 92–107. doi:10.2174/2213235X11301010092.
- [43] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics-A review in human disease diagnosis, *Anal. Chim. Acta.* 659 (2010) 23–33. doi:10.1016/j.aca.2009.11.042.
- [44] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg, F. Wulfert, Proposed minimum reporting standards for data analysis in metabolomics,

- Metabolomics. 3 (2007) 231–241. doi:10.1007/s11306-007-0081-3.
- [45] P.D. Whitfield, A.J. German, P.M. Noble, Metabolomics: an emerging post-genomic tool for nutrition, *Br. J. Nutr.* 92 (2004) 549–555. doi:10.1079/BJN20041243.
- [46] Handbook of ion chromatography, *Choice Rev. Online.* 42 (2005) 42-6508-42–6508. doi:10.5860/CHOICE.42-6508.
- [47] B. Meermann, V. Nischwitz, ICP-MS for the analysis at the nanoscale-a tutorial review, *J. Anal. At. Spectrom.* 33 (2018) 1432–1468. doi:10.1039/c8ja00037a.
- [48] Y.B. Monakhova, H. Schäfer, E. Humpfer, M. Spraul, T. Kuballa, D.W. Lachenmeier, Application of automated eightfold suppression of water and ethanol signals in ^1H NMR to provide sensitivity for analyzing alcoholic beverages, *Magn. Reson. Chem.* 49 (2011) 734–739. doi:10.1002/mrc.2823.
- [49] A.H. Emwas, E. Saccenti, X. Gao, R.T. McKay, V.A.P.M. dos Santos, R. Roy, D.S. Wishart, Recommended strategies for spectral processing and post-processing of 1D ^1H -NMR data of biofluids with a particular focus on urine, *Metabolomics.* 14 (2018) 1–23. doi:10.1007/s11306-018-1321-4.
- [50] AnalyzerPro Software WebPage, (2019). <http://www.spectralworks.com/products/analyzerpro/> (accessed March, 2020).
- [51] AMDI Software WebPage, (2019). <http://chemdata.nist.gov/dokuwiki/doku.php?id%4chemdata:amdi>



s (accessed March, 2020).

- [52] ChromaTOF Software WebPage, (2019). <https://www.leco.com/product/chromatof-software> (accessed March, 2020).
- [53] DeconvolutionReporting Software WebPage, (2019). [https://www.agilent.com/en/products/software-informatics/massspec-workstations/deconvolution-reporting-software-\(drs\)](https://www.agilent.com/en/products/software-informatics/massspec-workstations/deconvolution-reporting-software-(drs)) (accessed March, 2020).
- [54] MetaboliteDetector Software WebPage, (2019). <https://metabolitedetector.tu-bs.de/> (accessed March, 2020).
- [55] MS-DIAL Software WebPage, (2019). http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/ (accessed March, 2020).
- [56] MzMine Software WebPage, (2019). <http://mzmine.github.io/> (accessed March, 2020).
- [57] TraceFinder Software WebPage, (2019). <https://www.thermofisher.com/es/es/home/products-and-services/promotions/industrial/download-tracefinder-free-trial.html> (accessed March, 2020).
- [58] E. Hatzakis, Nuclear Magnetic Resonance (NMR) Spectroscopy in food science: a comprehensive review, *Compr. Rev. Food Sci. Food Saf.* 18 (2019) 189–220. doi:10.1111/1541-4337.12408.
- [59] D. Jacob, C. Deborde, M. Lefebvre, M. Maucourt, A. Moing, NMRProcFlow: a graphical and interactive tool dedicated to 1D

- spectra processing for NMR-based metabolomics, *Metabolomics*. 13 (2017) 1–5. doi:10.1007/s11306-017-1178-y.
- [60] R.A. van den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: Improving the biological information content of metabolomics data, *BMC Genomics*. 7 (2006) 1–15. doi:10.1186/1471-2164-7-142.
- [61] Compound Discoverer Software, (2019). <https://www.thermofisher.com/es/es/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/multi-omics-data-analysis/compound-discoverer-software.html> (accessed March, 2020).
- [62] MATLAB Software, 2019. (n.d.). <https://es.mathworks.com/products/matlab.html> (accessed March, 2020).
- [63] Metaboanalyst Software, (2019). <http://www.metaboanalyst.ca> (accessed March, 2020).
- [64] MetabolAnalyze Software, (2019). https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjWmZvKw_jlAhWwyoUKHbUWAMEQFjAAegQIBhAB&url=https%3A%2F%2Fcran.r-project.org%2Fpackage%3DMetabolAnalyze&usg=AOvVaw3gZuDDURZ8xA8-ShMjx4Cl (accessed March, 2020).
- [65] SIMCA Software, (2019). <https://umetrics.com/products/simca> (accessed March, 2020).



- [66] Unscrambler Software, (2019).
<https://www.camo.com/unscrambler/> (accessed March, 2020).
- [67] Workflow4metabolomics Software, (2019).
<http://workflow4metabolomics.org> (accessed March, 2020).
- [68] XCMX Online Software, (2019).
https://xcmonline.scripps.edu/landing_page.php?pgcontent=mainPage (accessed March, 2020).
- [69] J. Bouhlef, D.J. Bouveresse, S. Abouelkaram, E. Baéza, C. Jondreville, A. Travel, J. Ratel, E. Engel, D.N. Rutledge, Comparison of common components analysis with principal components analysis and independent components analysis: Application to SPME-GC-MS volatolomic signatures, *Talanta*. 178 (2018) 854–863. doi:10.1016/j.talanta.2017.10.025.
- [70] D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective, *Trends Food Sci. Technol.* 72 (2018) 83–90. doi:10.1016/j.tifs.2017.12.006.
- [71] Å.M. Wheelock, C.E. Wheelock, Trials and tribulations of ‘omics data analysis: assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine, *Mol. Biosyst.* 9 (2013) 2589. doi:10.1039/c3mb70194h.
- [72] L.C. Lee, C.Y. Liang, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-

- dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps, *Analyst*. 143 (2018) 3526–3539. doi:10.1039/c8an00599k.
- [73] A. Kalivodová, K. Hron, P. Filzmoser, L. Najdekr, H. Janečková, T. Adam, PLS-DA for compositional data with application to metabolomics, *J. Chemom.* 29 (2015) 21–28. doi:10.1002/cem.2657.
- [74] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment - A review, *Anal. Chim. Acta*. 891 (2015) 1–14. doi:10.1016/j.aca.2015.04.042.
- [75] T. Schönberger, Y.B. Monakhova, D.W. Lachenmeier, S. Walch, T. Kuballa, N.-P.E. Team, (NEXT) -NMR working group Germany, EUROLAB Technical Report 01/2015 -Guide to NMR Method Development and Validation-Part II: Multivariate data analysis, (2015) 1–20. [http://www.eurolab.org/documents/NMR Val Guideline II V6.pdf](http://www.eurolab.org/documents/NMR_Val_Guideline_II_V6.pdf).
- [76] M.N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D.N. Rutledge, P. Savarin, PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters, *Mol. BioSyst.* 11 (2015) 13–19. doi:10.1039/C4MB00414K.
- [77] R. Bünger, R.T. Mallet, Metabolomics and Receiver Operating Characteristic Analysis, *Crit. Care Med.* 44 (2016) 1784–1785. doi:10.1097/CCM.0000000000001795.
- [78] W. Chen, Y. Du, F. Zhang, R. Zhang, B. Ding, Z. Chen, Q. Xiong,



- Sampling error profile analysis (SEPA) for model optimization and model evaluation in multivariate calibration, *J. Chemom.* 32 (2018) e2933. doi:10.1002/cem.2933.
- [79] F. Tosato, R.M. Correia, B.G. Oliveira, A. Mazega, H. Franca, W.K.T. Coltro, P.R. Filgueiras, W. Romão, Paper spray ionization mass spectrometry allied to chemometric tools for quantification of whisky adulteration with additions of sugar cane spirit., *Anal. Methods.* 10 (2018). doi:10.1039/C8AY00071A.
- [80] B. Galindo-Prieto, L. Eriksson, J. Trygg, Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis, *Chemom. Intell. Lab. Syst.* 146 (2015) 297–304. doi:10.1016/j.chemolab.2015.05.001.
- [81] ChEBI, (2019). <https://www.ebi.ac.uk/chebi/> (accessed March, 2020).
- [82] ChemSpider, (2019). <http://www.chemspider.com/> (accessed March, 2020).
- [83] HMDB, (2019). <http://www.hmdb.ca/> (accessed March, 2020).
- [84] KEGG, (2019). <https://www.genome.jp/kegg/> (accessed March, 2020).
- [85] NIST, (2019). <https://webbook.nist.gov/chemistry/> (accessed March, 2020).
- [86] LipidMaps, (2019). <https://www.lipidmaps.org/> (accessed March, 2020).

- [87] MassBank, (2019). <http://www.massbank.jp/> (accessed March, 2020).
- [88] METLIN, (2019). https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage (accessed March, 2020).
- [89] PubChem, (2019). <https://pubchem.ncbi.nlm.nih.gov/> (accessed March, 2020).
- [90] K. Bingol, L. Bruschweiler-Li, C. Yu, A. Somogyi, F. Zhang, R. Brüschweiler, Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures, *Anal. Chem.* 87 (2015) 3864–3870. doi:10.1021/ac504633z.
- [91] A.C. Dona, M. Kyriakides, F. Scott, E.A. Shephard, D. Varshavi, K. Veselkov, J.R. Everett, A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments, *Comput. Struct. Biotechnol. J.* 14 (2016) 135–153. doi:10.1016/j.csbj.2016.02.005.
- [92] J.L. Markley, R. Brüschweiler, A.S. Edison, H.R. Eghbalnia, R. Powers, D. Raftery, D.S. Wishart, The future of NMR-based metabolomics, *Curr. Opin. Biotechnol.* 43 (2017) 34–40. doi:10.1016/j.copbio.2016.08.001.
- [93] A. Smolinska, J. Engel, E. Szymanska, L. Buydens, L. Blanchet, General framing of low-, mid-, and high-level data fusion with examples in the life sciences, in: *Data Handl. Sci. Technol.*, 2019: pp. 51–79. doi:10.1016/B978-0-444-63984-4.00003-X.



- [94] S. Dai, Z. Lin, B. Xu, Y. Wang, X. Shi, Y. Qiao, J. Zhang, Metabolomics data fusion between near infrared spectroscopy and high-resolution mass spectrometry: A synergetic approach to boost performance or induce confusion, *Talanta*. 189 (2018) 641–648. doi:10.1016/j.talanta.2018.07.030.
- [95] M. Cocchi, Introduction: Ways and Means to Deal With Data From Multiple Sources, *Data Handl. Sci. Technol.* 31 (2019) 1–26. doi:10.1016/B978-0-444-63984-4.00001-6.
- [96] R. Ríos-Reina, R.M. Callejón, F. Savorani, J.M. Amigo, M. Cocchi, Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars, *Talanta*. 198 (2019) 560–572. doi:10.1016/j.talanta.2019.01.100.
- [97] A. Biancolillo, R. Bucci, A.L. Magrì, A.D. Magrì, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication, *Anal. Chim. Acta*. 820 (2014) 23–31.
- [98] E. Borràs, J. Ferré, R. Boqué, L. Aceña, A. Calvo, O. Busto, Prediction of olive oil sensory descriptors using instrumental data fusion and partial least squares (PLS) regression, *Talanta*. 155 (2016) 116–123. doi:10.1016/j.talanta.2016.04.040.
- [99] V. Mlynárik, Introduction to nuclear magnetic resonance, *Anal. Biochem.* 529 (2017) 4–9. doi:10.1016/j.ab.2016.05.006.
- [100] T.C. Pochapsky, S.S. Pochapsky, Nuclear magnetic resonance spectroscopy, in: *Mol. Biophys. Life Sci.*, 2013. doi:10.1007/978-1-4614-8548-3_5.

- [101] B.L. Marquez, R.T. Williamson, Quantitative applications of NMR spectroscopy, *Chem. Eng. Pharm. Ind. R&D to Manuf.* (2010) 581–596. doi:10.1002/9780470882221.ch31.
- [102] T. Huang, H. Li, W. Zhang, M. Numata, L. Mackay, J. Warren, H. Jiao, S.W. Westwood, D. Song, Advanced approaches and applications of qNMR, *Metrologia.* (2019) 0–8. doi:10.1088/1681-7575/ab336b.
- [103] C. de Souza Araújo, A.P. de Oliveira, A.D. da Conceição Santos, A.L. Guimarães, N.D. dos Santos Silva, M.A.Á. Queiroz, E.C. da Cruz Araújo, J.R.G. da Silva Almeida, Total content of kaurene diterpenes in *Annona vepretorum* stems via ^1H qNMR: A method for speeding the identification of bioactive extracts, *Phytochem. Anal.* 30 (2019) 83–88. doi:10.1002/pca.2792.
- [104] D. Guimarães, J. Noro, A. Loureiro, A. Cavaco-Paulo, E. Nogueira, Quantification of drugs encapsulated in liposomes by ^1H NMR, *Colloids Surfaces B Biointerfaces.* 179 (2019) 414–420. doi:10.1016/j.colsurfb.2019.03.039.
- [105] P.H.C. Franco, S.F.P. Braga, R.B. de Oliveira, I.C. César, Purity determination of a new antifungal drug candidate using quantitative ^1H NMR spectroscopy: Method validation and comparison of calibration approaches, *Magn. Reson. Chem.* (2019) 1–9. doi:10.1002/mrc.4936.
- [106] L. López-Garrido, A.E. Bañuelos-Hernández, E. Pérez-Hernández, R. Tecualt-Gómez, J. Quiroz-Williams, A. Ariza-Castolo, E. Becerra-Martínez, N. Pérez-Hernández, Metabolic profiling of serum in patients with cartilage tumours using ^1H -



- NMR spectroscopy: A pilot study , *Magn. Reson. Chem.* (2019) 1–12. doi:10.1002/mrc.4925.
- [107] D.P. Hollis, Quantitative Analysis of Aspirin, Phenacetin, and Caffeine mixtures by nuclear magnetic resonance spectrometry, *Anal. Chem.* 35 (1963) 1682–1684. doi:10.1021/ac60204a043.
- [108] B. Diehl, U. Holzgrabe, Y. Monakhova, T. Schönberger, Quo Vadis qNMR?, *J. Pharm. Biomed. Anal.* (2019) 112847. doi:10.1016/j.jpba.2019.112847.
- [109] P. Giraudeau, Challenges and perspectives in quantitative NMR, *Magn. Reson. Chem.* 55 (2017) 61–69. doi:10.1002/mrc.4475.
- [110] R.A.F. Cavalcante, F.L. Silva, F. Favero, I.S. Resck, A.L. Pereira, A.H.L. Machado, Quantitative ^1H NMR spectroscopy (qNMR) in the early process development of a new quorum sensing inhibitor , *Magn. Reson. Chem.* (2019) 1–10. doi:10.1002/mrc.4906.
- [111] Scopus, (n.d.). www.scopus.com (accessed March, 2020).
- [112] H. Friebolin, Basic One- and Two-Dimensional NMR Spectroscopy, 4th Edition, WILEY-VCH. (2013). doi:10.1055/s-2005-867116.
- [113] W. Levason, D. Pletcher, ^{195}Pt -Platinum Nuclear Magnetic Resonance spectroscopy: monitoring platinum electroplating baths, *Platin. Met. Rev.* 37 (1993) 17–23. <http://www.platinummetalsreview.com/pdf/pmr-v37-i1-017-023.pdf>.

- [114] T. Nozawa, M. Hatano, M. Sato, Y. Toida, E. Batholdi, 57 Fe Spin-lattice Relaxation Time of Fe(CO)₅ and Ferrocene, *Bull. Chem. Soc. Jpn.* 56 (1983) 3837–3838. doi:10.1246/bcsj.56.3837.
- [115] Z. Zhou, Y. He, X. Qiu, D. Redwine, J. Potter, R. Cong, M. Miller, Optimum Cr(acac)₃ concentration for NMR quantitative analysis of polyolefins, *Macromol. Symp.* 330 (2013) 115–122. doi:10.1002/masy.201300034.
- [116] M. Weber, C. Hellriegel, A. Rueck, J. Wuethrich, P. Jenks, M. Obkircher, Method development in quantitative NMR towards metrologically traceable organic certified reference materials used as ³¹P qNMR standards, *Anal. Bioanal. Chem.* 407 (2015) 3115–3123. doi:10.1007/s00216-014-8306-6.
- [117] S. Demkowicz, J. Rachon, M. Daśko, W. Kozak, Selected organophosphorus compounds with biological activity. Applications in medicine, *RSC Adv.* 6 (2016) 7101–7112. doi:10.1039/c5ra25446a.
- [118] M. Edén, Recent progress in homonuclear correlation spectroscopy of quadrupolar nuclei, in: *Mod. Magn. Reson.*, 2018. doi:10.1007/978-3-319-28388-3_104.
- [119] R. Youngman, NMR spectroscopy in glass science: A review of the elements, *Materials* (Basel). 11 (2018). doi:10.3390/ma11040476.
- [120] P. Bendel, Biomedical applications of ¹⁰B and ¹¹B NMR, *NMR Biomed.* 18 (2005) 74–82. doi:10.1002/nbm.886.
- [121] E. Hey-Hawkins, C.V. Teixidor, eds., *Boron-Based Compounds*,



John Wiley & Sons, Ltd, Chichester, UK, 2018.
doi:10.1002/9781119275602.

- [122] H. Maki, G. Sakata, M. Mizuhata, Quantitative NMR of quadrupolar nucleus as a novel analytical method: hydrolysis behaviour analysis of aluminum ion †, *Analyst*. 142 (2017) 1790–1799. doi:10.1039/c7an00067g.
- [123] T.D.W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry: Third Edition*, 2016. doi:10.1016/C2015-0-04654-8.
- [124] E. Ralli, M. Amargianitaki, E. Manolopoulou, M. Misiak, G. Markakis, S. Tachtalidou, A. Kolesnikova, P. Dais, A. Spyros, NMR Spectroscopy protocols for food metabolomics applications, in: *Metab. Profiling Methods Protoc. Methods Mol. Biol.*, 2018: pp. 203–211. doi:10.1007/978-1-4939-7643-0_14.
- [125] G.F. Pauli, S.N. Chen, C. Simmler, D.C. Lankin, T. Godecke, B.U. Jaki, J.B. Friesen, J.B. McAlpine, J.G. Napolitano, Importance of purity evaluation and the potential of quantitative ¹H NMR as a purity assay, *J. Med. Chem.* 57 (2014) 9220–9231. doi:10.1021/jm500734a.
- [126] R.K. Harris, E.D. Becker, S.M. Cabral De Menezes, P. Granger, R.E. Hoffman, K.W. Zilm, Further Conventions for NMR Shielding and Chemical Shifts (IUPAC Recommendations 2008) Copyright IUPAC (2008). Reprinted with permission from *Pure Appl. Chem.*, 2008, 80, 59-84, which can be freely downloaded from www.iupac.org/publications/pac/80/1/0059, p, *Enycl. Magn. Reson.* (2008).

doi:10.1002/9780470034590.emrstm1019.

- [127] R.K. Harris, E.D. Becker, S.M. Cabral de Menezes, R. Goodfellow, P. Granger, NMR Nomenclature: nuclear spin properties and conventions for chemical shifts, *Solid State Nucl. Magn. Reson.* 22 (2002) 458–483. doi:10.1006/snmr.2002.0063.
- [128] C. Yu, Q. Zhang, P. Xu, Y. Bai, W. Shen, B. Di, M. Su, Quantitative determination and validation of octreotide acetate using ^1H -NMR spectroscopy with internal standard method, *Magn. Reson. Chem.* (2017) 37–45. doi:10.1002/mrc.4665.
- [129] L. Brennan, NMR-based metabolomics: From sample preparation to applications in nutrition research, *Prog. Nucl. Magn. Reson. Spectrosc.* (2014). doi:10.1016/j.pnmrs.2014.09.001.
- [130] Y.B. Monakhova, B.W.K. Diehl, Monitoring daily routine performance in quantitative NMR (qNMR) spectroscopy: Is the system suitability test necessary?, *Magn. Reson. Chem.* 57 (2019) 110–117. doi:10.1002/mrc.4798.
- [131] V. Gallo, N. Intini, P. Mastrorilli, M. Latronico, P. Scapicchio, M. Triggiani, V. Bevilacqua, P. Fanizzi, D. Acquotti, C. Airoidi, F. Arnesano, M. Assfalg, F. Benevelli, D. Bertelli, L.R. Cagliani, L. Casadei, F. Cesare Marincola, G. Colafemmina, R. Consonni, C. Cosentino, S. Davalli, S.A. De Pascali, V. D’Aiuto, A. Faccini, R. Gobetto, R. Lamanna, F. Liguori, F. Longobardi, D. Mallamace, P. Mazzei, I. Menegazzo, S. Milone, A. Mucci, C. Napoli, T. Pertinhez, A. Rizzuti, L. Rocchigiani, E. Schievano, F. Sciubba, A. Sobolev, L. Tenori, M. Valerio, Performance Assessment in Fingerprinting and Multi Component Quantitative NMR



- Analyses, *Anal. Chem.* 87 (2015) 6709–6717.
doi:10.1021/acs.analchem.5b00919.
- [132] C. Chartrand-Lefebvre, M. Durand, C. Tremblay, Comisión Europea, Decisión de la comisión 2002/657/CE por laCAPÍTULO 1. INTRODUCCIÓN 77 que se aplica la Directiva 96/23/CE del Consejo en cuanto al funcionamiento de los métodos analíticos y la interpretación de los resultados, 2014. doi:10.7326/L14-5033.
- [133] The European Commission, Sante/11813/2017 Analytical Quality Control and Method Validation Procedures for Pesticide Residues Analysis, (2018) 1–46. doi:10.13140/RG.2.2.33021.77283.
- [134] P.P. Morillas, Guía Eurachem: La adecuación al uso de los métodos analíticos – Una Guía de laboratorio para la validación de métodos y temas relacionados, 1ª ed., Eurolab España, n.d. www.eurachem.org.
- [135] FDA, Guidelines for the Validation of Chemical Methods for the FDA Foods Program, *Foods Progr. Sci. Res. Steer. Comm.* (2015) 37.
<https://www.fda.gov/downloads/scienceresearch/fieldscience/ucm273418.pdf>.
- [136] J. et al. Ammon, Guide to NMR Method Development and Validation – Part 1: Identification and Quantification, EUROLAB Tech. Rep. 1/2014. (2014). www.eurolab.org (accessed March, 2020).
- [137] M. Halabalaki, K. Vougiannopoulou, E. Mikros, A.L. Skaltsounis, Recent advances and new strategies in the NMR-

- based identification of natural products, *Curr. Opin. Biotechnol.* 25 (2014) 1–7. doi:10.1016/j.copbio.2013.08.005.
- [138] M. Elyashberg, K. Blinov, S. Molodtsov, A.J. Williams, Structure Revision of Asperjinone Using Computer-Assisted Structure Elucidation Methods, *J. Nat. Prod.* 76 (2013) 113–116. doi:10.1021/np300218g.
- [139] J.B. McAlpine, S.-N. Chen, A. Kutateladze, J.B. MacMillan, G. Appendino, A. Barison, M.A. Beniddir, M.W. Biavatti, S. Bluml, A. Boufridi, M.S. Butler, R.J. Capon, Y.H. Choi, D. Coppage, P. Crews, M.T. Crimmins, M. Csete, P. Dewapriya, J.M. Egan, M.J. Garson, G. Genta-Jouve, W.H. Gerwick, H. Gross, M.K. Harper, P. Hermanto, J.M. Hook, L. Hunter, D. Jeannerat, N.-Y. Ji, T.A. Johnson, D.G.I. Kingston, H. Koshino, H.-W. Lee, G. Lewin, J. Li, R.G. Linington, M. Liu, K.L. McPhail, T.F. Molinski, B.S. Moore, J.-W. Nam, R.P. Neupane, M. Niemitz, J.-M. Nuzillard, N.H. Oberlies, F.M.M. Ocampos, G. Pan, R.J. Quinn, D.S. Reddy, J.-H. Renault, J. Rivera-Chávez, W. Robien, C.M. Saunders, T.J. Schmidt, C. Seger, B. Shen, C. Steinbeck, H. Stuppner, S. Sturm, O. Tagliatalata-Scafati, D.J. Tantillo, R. Verpoorte, B.-G. Wang, C.M. Williams, P.G. Williams, J. Wist, J.-M. Yue, C. Zhang, Z. Xu, C. Simmler, D.C. Lankin, J. Bisson, G.F. Pauli, The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research, *Nat. Prod. Rep.* 36 (2019) 35–107. doi:10.1039/C7NP00064B.
- [140] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R.



- Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0—The Human Metabolome Database in 2013, *Nucleic Acids Res.* 41 (2012) D801–D807. doi:10.1093/nar/gks1065.
- [141] N.S. Kale, K. Haug, P. Conesa, K. Jayseelan, P. Moreno, P. Rocca-Serra, V.C. Nainala, R.A. Spicer, M. Williams, X. Li, R.M. Salek, J.L. Griffin, C. Steinbeck, MetaboLights: An Open-Access Database Repository for Metabolomics Data, in: *Curr. Protoc. Bioinforma.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016; pp. 14.13.1-14.13.18. doi:10.1002/0471250953.bi1413s53.
- [142] J. Xia, I. V. Sinelnikov, B. Han, D.S. Wishart, MetaboAnalyst 3.0—making metabolomics more meaningful, *Nucleic Acids Res.* 43 (2015) W251–W257. doi:10.1093/nar/gkv380.
- [143] M. Foroozandeh, L. Castañar, L.G. Martins, D. Sinnaeve, G.D. Poggetto, C.F. Tormena, R.W. Adams, G.A. Morris, M. Nilsson, NMR Spectroscopy Hot Paper Ultrahigh-Resolution Diffusion-Ordered Spectroscopy *Zuschriften*, (2016) 15808–15811. doi:10.1002/ange.201609676.

CAPÍTULO PRIMERO. MARCO TEÓRICO



2 CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON

2.1 Introducción

Hoy en día el control de calidad en bebidas alcohólicas tanto para la comprobación de materias primas como de los productos finales es primordial para su caracterización óptima a fin de garantizar sus principales propiedades organolépticas y la seguridad alimentaria.

Esta importancia es debida a que la mayoría de los procesos de elaboración con frecuencia son de tipo artesanal, lo que les confiere unas características muy concretas y únicas en función del producto. Es el caso de las bebidas alcohólicas, donde su elaboración basada en la experiencia de los maestros artesanos podría ser más fiable y robusta con la aplicación de técnicas analíticas combinadas con herramientas de la metabolómica.

Un claro ejemplo es el ron, una de las bebidas alcohólicas obtenidas mediante destilación más consumidas en el mundo, concretamente el ron dorado. Se trata de una bebida espirituosa obtenida de la fermentación alcohólica y la destilación de melaza o jarabes de caña de azúcar. A tal fin, el proceso de preparación del mismo consta de distintas etapas: selección de la materia prima, fermentación, destilación, dilución, envejecimiento y mezclado (Figura 6).

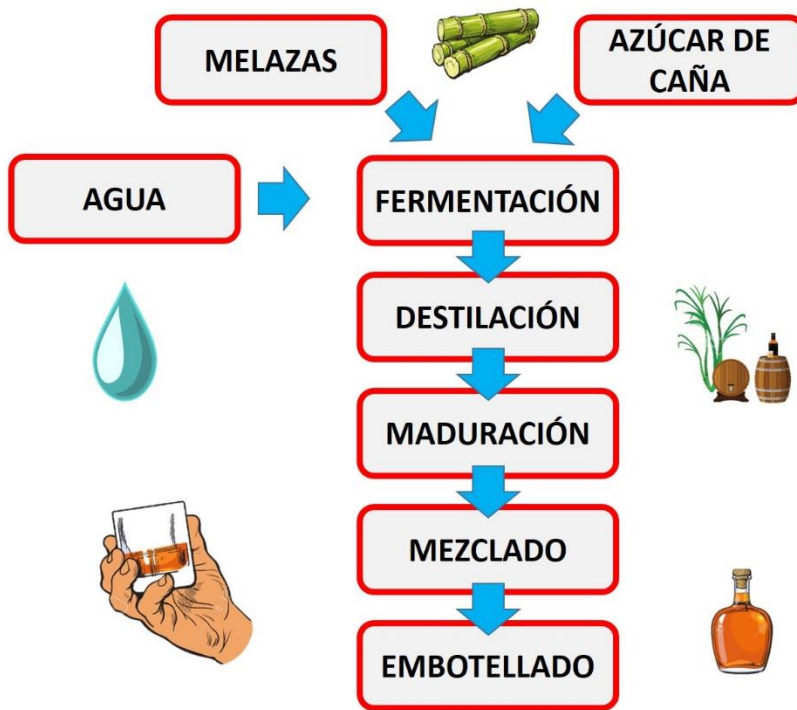


Figura 6. Diagrama de flujo de elaboración de ron

Además, la consideración de aquellas bebidas de ron de más alta calidad (como aquellos considerados *Premium*) les permite alcanzar un mayor precio de mercado, propiciando sus adulteraciones con fines ilícitos y lucrativos. También existe cierta controversia en la legislación vigente respecto al etiquetado. Según la legislación de la Unión Europea y de los Estados Unidos, la declaración de edad en la etiqueta debe referirse al ron más joven en la botella, dado que en general los productos comerciales son mezclas de rones de varias edades. Sin embargo, en otros países, puede referirse al ron más antiguo. Por último, la gran variabilidad que con frecuencia existe en todos los procesos de preparación, muestran la necesidad de un control adecuado desde el



punto de vista químico-analítico, con métodos tales como el que se propone en el presente trabajo. Por todo ello, el desarrollo de métodos analíticos fiables para el control de la calidad de estos productos resulta de gran interés ante la poca o escasa normativa vigente en cuanto a parámetros e indicadores de calidad tanto a nivel nacional como en la Unión Europea.

Los trabajos de ron publicados a día de hoy se encuentran muy limitados a estudios específicos de sus características. La similitud entre ron y su homólogo brasileño *cachaça* ha sido estudiada en base a las diferencias en su producción que dan lugar a propiedades organolépticas diferentes [1]. Estudios para la evaluación del aroma del ron también han aportado resultados interesantes, encontrándose compuestos como el *cis*-whiskey lactona, vanillina, ácido decanoico y 2-3-metilbutanol como compuestos mayoritarios responsables del olor característico [2,3]. Por otro lado, se han realizado trabajos de clasificación de carácter binario para la diferenciación entre melazas y azúcar de caña, con un 88% de acierto [6], o en función del origen (clasificando rones cubanos frente al resto) y contando también con resultados cercanos al 90% de acierto [4]. La Tabla 10 muestra una revisión bibliográfica con los artículos más recientes y destacados acerca de estudios de ron. La escasez de estudios enfocados a la evaluación general en términos de control y calidad del ron, así como la carencia de validaciones estrictas de los modelos clasificatorios, demuestran la necesidad de estudio del mismo.

CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON

Tabla 10. Revisión bibliográfica sobre estudios de ron

Título de la publicación	Referencia
<i>Cachaça and rum</i>	[1]
<i>Characterisation of odour-active compounds in aged rum</i>	[2]
<i>Characterization of the key aroma compounds in two commercial rums by means of the sensomics approach</i>	[3]
<i>Chemical profile of rums as a function of their origin. The use of chemometric techniques for their identification</i>	[4]
<i>Differentiation of rum and brazilian artisan cachaca via electrospray ionization mass spectrometry fingerprinting</i>	[5]
<i>Differentiation of rums produced from sugar cane juice (rhum agricole) from rums manufactured from sugar cane melazas by a metabolomics approach</i>	[6]
<i>Influence of the production process on the key aroma compounds of rum: from melazas to the spirit</i>	[7]

Por todo ello, en este capítulo se han abordado los siguientes objetivos. Por un lado, se ha desarrollado una estrategia para la clasificación de ron a partir de su fracción volátil y/o semivolátil usando herramientas de la metabolómica a través de huellas dactilares adquiridas por SPME por espacio en cabeza (HS) acoplada a GC-MS. Por otro lado, se ha llevado a cabo un estudio metabolómico de ron dorado mediante la adquisición de huellas dactilares de sus componentes no volátiles tras extracción tipo *dilute and shoot* y análisis LC-HRMS. Ambos estudios han permitido la identificación de marcadores químicos de calidad del ron. Asimismo, se ha procedido a la adquisición de huellas dactilares a través de $^1\text{H-NMR}$ para el estudio de los compuestos mayoritarios presentes en rones dorados comerciales. Finalmente se ha aplicado la fusión de datos multi-técnica (GC-MS, LC-MS y NMR) a fin de comparar las capacidades discriminatorias de las diferentes técnicas y su fusión en los análisis multivariantes aplicados.



Las contribuciones científicas correspondientes a este capítulo son:

- **Aportación Científica 1.** *Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry*
- **Aportación Científica 2.** *An Innovative Metabolomic Approach for Golden Rum Classification Combining Ultrahigh-Performance Liquid Chromatography–Orbitrap Mass Spectrometry and Chemometric Strategies*
- **Aportación Científica 3.** *$^1\text{H-NMR}$ and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis*

2.2 *Bibliografía*

- [1] A.B.P. Medeiros, M.E. de Matos, A. de Pinho Monteiro, J.C. de Carvalho, C.R. Soccol, Cachaça and rum, in: *Curr. Dev. Biotechnol. Bioeng.*, Elsevier, 2017: pp. 451–468. doi:10.1016/B978-0-444-63666-9.00016-9.
- [2] J.A. Pino, S. Tolle, R. Gök, P. Winterhalter, Characterisation of odour-active compounds in aged rum, *Food Chem.* 132 (2012) 1436–1441. doi:10.1016/j.foodchem.2011.11.133.
- [3] L. Franitza, M. Granvogl, P. Schieberle, Characterization of the key aroma compounds in two commercial rums by means of the sensomics approach, *J. Agric. Food Chem.* 64 (2016) 637–645. doi:10.1021/acs.jafc.5b05426.
- [4] O.M. Sampaio, R.V. Reche, D.W. Franco, Chemical profile of rums as a function of their origin. The use of chemometric techniques for their identification, *J. Agric. Food Chem.* 56 (2008) 1661–1668. doi:10.1021/jf0726841.
- [5] P.P. de Souza, D. V Augusti, R.R. Catharino, H.G.L. Siebald, M.N. Eberlin, R. Augusti, Differentiation of rum and Brazilian artisan cachaca via electrospray ionization mass spectrometry fingerprinting, *J. Mass Spectrom.* 42 (2007) 1294–1299.
- [6] L. Franitza, L. Nicolotti, M. Granvogl, P. Schieberle, Differentiation of rums produced from sugar cane juice (Rhum Agricole) from rums manufactured from sugar cane molasses by a metabolomics approach, *J. Agric. Food Chem.* 66 (2018) 3038–3045. doi:10.1021/acs.jafc.8b00180.



- [7] L. Frantza, M. Granvogl, P. Schieberle, Influence of the production process on the key aroma compounds of rum: from molasses to the spirit, *J. Agric. Food Chem.* 64 (2016) 9041–9053. doi:10.1021/acs.jafc.6b04046.

CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON





2.3 Aportación científica 1. Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

J.R. Belmonte-Sánchez, S. Gherghel, J. Arrebola-Liébanas, R.R. González, J.L.M. Vidal, I. Parkin, A.G. Frenich. *Talanta*. 187 (2018) 348–356.

DOI: 10.1016/j.talanta.2018.05.025

CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON



Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

José Raúl Belmonte-Sánchez^a, Simona Gherghel^{a,b,c}, Javier Arrebola-Liébanas^a, Roberto Romero González^a, José Luis Martínez Vidal^a, Ivan Parkin^c, Antonia Garrido French^{a,*}

^a Research Group “Analytical Chemistry of Contaminants”, Department of Chemistry and Physics, Research Centre for Agricultural and Food Biotechnology (BITAL), University of Almería, Agrifood Campus of International Excellence, ceiA3, E-04120 Almería, Spain

^b UCL Department of Security and Crime Science, 35 Tavistock Square, London WC1H 9EZ, United Kingdom

^c UCL Department of Chemistry, 20 Gordon Street, London WC1H 0AJ, United Kingdom

ARTICLE INFO

Keywords:

Rum
Classification
Volatile organic compounds
Multivariate analysis
SPME-GC-MS

ABSTRACT

In this study, targeted and untargeted analyses based on headspace solid phase microextraction coupled to gas chromatography-mass spectrometry (HS-SPME-GC-MS) method were developed for classifying 33 different commercial rums. Targeted analysis showed correlation of ethyl acetate and ethyl esters of carboxylic acids with aging when rums of the same brand were studied, but presented certain limitations when the comparison was carried out between different brands. To overcome these limitations, untargeted strategies based on unsupervised treatments, such as hierarchical cluster analysis (HCA) and principal component analysis (PCA), as well as supervised methods, such as linear discriminant analysis (LDA) were applied. HCA allowed distinguishing main groups (with and without additives), while the PCA method indicated 40 ions corresponding to 13 discriminant compounds as relevant chemical descriptors for the correct rum classification (PCA variance of 88%). The compounds were confirmed based on the combination of retention indexes and low and high-resolution mass spectrometry (HRMS). Using the obtained results, LDA was carried out for the analytical discrimination of the remaining rums based on manufacturing country, raw material type, distillation method, wood barrel type and aging period and 94%, 91%, 92%, 95% and 94% of rums, respectively, were correctly classified. The proposed methodology has led to a robust analytical strategy for the classification of rums as a function of different parameters depending on the rum production process.

1. Introduction

Rum is a fairly aromatic spirit, obtained exclusively from sugar cane juice or molasses, and then subjected to the processes of alcoholic fermentation, distillation and aging. This spirit represents a widely popular alcoholic beverage with a high world consumption rate (more than 1 billion of litres per year) and an expected increase of 1.9% in volume terms over 2016–2021. [1,2].

The complex elaboration of this type of alcoholic beverage makes it an attractive object of study. Differences in the production process are known to lead to wide variability in its composition, although this variation has not been fully understood yet [3,4]. The production process begins with the fermentation of the chosen raw material, which leads to the formation of a number of volatile compounds, such as alcohols, ethyl esters and aldehydes, among others [5]. The resulting

mash is distilled using heat in copper pot stills or in stainless steel columns to obtain a high content of ethanol, which inevitably leads to the loss of some aroma compounds [6,7]. Additionally, different distillation methods can be applied, such as continuous and batch distillation (e.g. Jamaican “heavy rums” typically made by batch distillation) [8]. The resulting distillate is diluted with pure demineralized water to obtain an alcohol percentage of around 35–40%, which is then aged in oak barrels previously used for whiskey or brandy production [9,10]. The aging step gives rum its characteristic flavor as a large number of new compounds emerge. Ethyl esters are generated as a result of the high percentage of ethanol, while a number of different compounds such as whiskey lactone, vanillin and 2-methoxyphenol can form because of the interaction with the wood barrels [3]. Additionally, as rum matures, it generally gains golden hues as a result of the tannins from the barrel staves [11]. After an aging period, typically of at least 1

* Corresponding author.

E-mail address: agarrido@ual.es (A. Garrido French).

Table 1
Code, bottle labelling and manufacturer information for the rum samples analysed.

Brand code	Bottle code	Origin	Aging	Raw material	Distillation method	Barrel	
A	A_5Y	Cuba	5 years	Molasses	Stainless steel columns	Oak barrel	
	A_7Y		7 years	Molasses	Stainless steel columns	Oak barrel	
B	B_SYRUP1	Dominican Republic	B_1	Anejo	N/A	American oak barrel	
			B_2	Young	Molasses	N/A	American oak barrel
			B_3	N/A	Molasses	N/A	American oak barrel
C	C_10Y	Dominican Republic	C_7Y	7 years	Molasses	Stainless steel columns	
			C_10Y	10 years	Molasses	Stainless steel columns	American oak barrel
			C_15Y	15 years	Molasses	Stainless steel columns	American oak barrel
D	D_HONEY1	Dominican Republic	D	Anejo	N/A	N/A	
			D_R2	Anejo	N/A	N/A	
			D_HONEY1	N/A	N/A	N/A	
E	E		Old reserve	N/A	N/A	N/A	
F	F		Anejo	N/A	N/A	N/A	
G	G_1	Grenada	Double aged. Old reserve	Molasses	N/A	Mixed (American oak barrel + French oak barrel)	
			G_2	Trinidad & Tobago	Double aged. Overproof	Molasses	Stainless steel columns
H	H_18Y	Guatemala	H_8Y	8 years	Sugarcane juice	Stainless steel columns	
			H_12Y	12 years	Sugarcane juice	Stainless steel columns	American oak barrel
			H_18Y	18 years	Sugarcane juice	Stainless steel columns	American oak barrel
I	I_12Y	Jamaica	I_5Y	5–10 years	Molasses	Copper pot stills	
			I_12Y	12 years	Molasses	Copper pot stills	American oak barrel
J	J_18Y	Nicaragua	J_7Y	7 years	Molasses	Stainless steel columns	
			J_12Y	12 years	Molasses	Stainless steel columns	American oak barrel
			J_18Y	18 years	Molasses	Stainless steel columns	American oak barrel
K	K_VAN1	Republic of Mauritius	K_MAND1	Double aged	Sugarcane juice	Stainless steel columns	
			K_VAN1	Double aged	Sugarcane juice	Stainless steel columns	N/A
L	L_HONEY1	Spain	N/A	N/A	N/A	N/A	
M	M_R2		M	Anejo	Molasses	N/A	
			M_R2		Molasses	N/A	N/A
N	N		Dorado	N/A	N/A	N/A	
P	P		Anejo	N/A	N/A	N/A	
O	O		N/A	N/A	N/A	N/A	
Q	Q_10Y	Venezuela	Q_3Y	3 years	Molasses	Mixed (Steel column + a bit of copper)	
			Q_10Y	10 years	Molasses	Mixed (Steel column + a bit of copper)	American oak barrel

1: additives added.

2: same rum type purchased at a different liquor store.

N/A; information not available.

(2014) mass spectra database.

At the beginning of the analysis, the column temperature was set to 35 °C, and the temperature was increased to 100 °C at a 4 °C/min rate, and then to 250 °C (hold 20 min) at a rate of 20 °C/min. The total run time was 43.75 min.

The QQQ mass spectrometer was operated in full scan mode. The temperatures of the transfer line, manifold, and ionization source were set to 280, 40, and 280 °C, respectively. The electron multiplier voltage was set to 1600 V (+200 V offset above the value obtained in the auto-tuning process). Mass peak widths set in the first and third quadrupole were of m/z 1.5 and 2.0, respectively. The analysis was carried out in the range of mass/charge ratios of m/z 50–400.

The volatile compounds of interest were identified against a commercial library (NIST14) and by the use of GC retention indices. Retention time of each volatile was converted to the Kovats retention index using C7–C40 n-alkanes as references and verified with those reported in the literature. A retention index window of ± 20 was applied to MS peak identification assignment [43].

2.5. HRMS-Q-Exactive analysis

As an additional confirmation method, a Q-Exactive-GC hybrid quadrupole Orbitrap mass spectrometer (Q-Exactive™, Thermo Fisher Scientific, Bremen, Germany) was used with the same chromatographic conditions as the GC-QQQ-MS analysis reported above. A VF-5 ms capillary column (30 m x 0.25 mm i.d. x 0.25 μ m film thickness) from Varian (Palo Alto, California, USA) was used for GC separation. The

positive electron ionization (EI) source was operated at 70 eV at a temperature of 200 °C with a transfer line temperature of 250 °C. High-resolution mode was operated at a 60000 full width at half maximum (FWHM) resolving power (m/z 207).

SPME extraction was operated using an identical fiber selected for GC-QQQ-MS analysis (polydimethylsiloxane, 100 μ m film thickness). Quantitation was performed using Xcalibur 4.1 and TraceFinder 4.1 software (Thermo Fisher Scientific, Les Ulis, France).

2.6. Pre-processing and data treatment

The GC–MS raw data (XMS file format) were converted to CDF file format using the Openchrom software (Armonk, NY, USA) [44]. Then, the dataset was pre-processed with MZmine 2.23 software (Norwood, MA, USA) [45]. The MZmine's "3D Viewer" tool offered a three-dimensional representation of the total ion chromatogram (TIC) in order to work with useful ranges of intensity, retention time and m/z ranges, (Fig. S-1, Supporting Information). The pre-processing treatment consisted of the optimisation of the following steps: (i) centroid mass detection, (ii) chromatogram builder, (iii) chromatographic deconvolution, (iv) alignment, (v) peak list, (vi) duplicate peak filter and (vii) gap filling. The results were stored as a CSV format file.

The data set was imported as a XML file using Excel, version 2013 (Redmond, WA, USA) and then, it was processed by normalising the ion intensities between 0 and 1: [value-minimum value]/[maximum value-minimum value].

2.7. Processing and multivariate data analysis

All statistical analyses in this study were conducted using the SPSS 23.0 software package (SPSS Inc, Chicago, IL, USA). For sample classification and discrimination, HCA, PCA and LDA were applied.

3. Results and discussion

3.1. HS-SPME-GC-MS method

A HS-SPME procedure was developed to carry out the analysis of the volatile/semivolatile compounds in rums. For that, a typical commercial 7-years old rum was used.

The 100 μm polydimethylsiloxane (PDMS) SPME fiber was selected as it has been extensively used for extracting volatile/semi-volatile compounds with a wide polarity range [46], some applications in alcoholic drinks such as wine [47] and rum [6,35].

For the GC-QqQ-MS analysis, the starting oven temperature was set to 35 °C to allow the elution of the most volatile compounds. In order to avoid the chromatographic co-elution of high concentrations of low molecular weight alcohols, such as ethanol and minor volatile compounds, each sample was monitored in full scan mode in the m/z range 50–400.

3.2. Analysis of targeted compounds

Targeted analysis was focused on the determination of some selected compounds traditionally present in rum samples and already reported as relevant compounds for the organoleptic properties of such spirit drinks [3,4,6]. For this purpose, ethyl acetate, furfural, 5-hydroxymethylfurfural, 2,5-furandicarboxaldehyde, vanillin, 3-methyl-1-butanol, whiskey lactone and the ethyl esters of the majoritarian carboxylic acids between C8 and C16 (individual and as sum of all them) were monitored and found in most of the studied rum samples. Their identification was carried out by comparison of their mass spectra with NIST mass spectrum library and considering the retention index tabulated in bibliography [43].

Six of the studied rum brands were available at different ages. For them, the peak areas of various discriminant compounds were plotted to check the content variation among the samples. The data for ethyl acetate and the sum of the even ethyl esters between C8 and C16 are shown in Fig. 1. The data obtained for the individual ethyl esters are provided in the Supporting Information Fig. S2. It can be observed that with a few exceptions, there is a trend for an increased peak area with rum aging within the same brand. For example, ethyl acetate and the sum of the ethyl esters were found at higher amounts in older rums (I, J, C and Q samples). However, these compounds content was not satisfactorily correlated with aging between samples of different brands and therefore, they cannot be used as proper chemical indicators of aging in rums. For example, in both plots it can be observed that some

rum samples (i.e. C_15Y) contained lower amounts of ethyl acetate or ethyl esters than younger rum samples of different brands (i.e., A_7Y, I_5Y, H_8Y and H_12Y).

An evaluation of the presence of the target compounds in the different rums did not provide a clear correlation between their presence in the studied samples and other factors such as origin, raw material, distillation type, barrel type, etc. Therefore, to overcome the limitations observed in the targeted analysis, further chemometrical tools were evaluated for untargeted analysis.

3.3. Pre-processing and dataset treatment

In the first stage, the raw data export to CDF format was visualized using the 3D Viewer tool of MZmine 2.23. Details of the MZmine steps and parameters are summarized in Table S-1 (Supporting Information).

Once the alignment was completed and before the “Gap Filling” process, the extraneous ions detected in all the samples (corresponding to SPME fiber, rubber septum, column bleeding, etc.) were substrated. For that purpose, a blank sample was analysed between each analysis of every type of commercial rum. This allowed the possibility of subtracting the ions detected by the software in these samples (known interfering masses), that could lead to errors in classification. These ions were mostly represented by cyclosiloxane signals from the SPME fiber (m/z 222, 296, 370), siloxanes from the sample vial septa (m/z 73, 207, 281), and phthalates from plasticizers contamination (m/z 149).

The resulting data from MZmine were exported as a CSV file containing a matrix giving the ion intensity values for each rum sample. In the time window from 1.20 to 30.00 min for a mass range m/z from 50 to 400, when 96 analyses (33 different rum samples with three replicates each, except 3 rum samples that were studied in duplicate) were performed, 231 variables (number of ions detected as the relevant ones by MZmine from the GC-MS data) were obtained. Thus, a total of 22176 data points were processed and used for multivariate analysis.

3.4. Unsupervised methods: Exploratory data analysis

An exploratory data analysis by HCA was carried out. The HCA was applied in order to reduce the dimensions of the data set by grouping relatively similar samples in one cluster and relatively dis-similar objects in another. Besides, the HCA enabled to check the repeatability of the analyses, detect anomalous values and discard any outlier replicates. At this step, four outlier replicates out of a total of 96 rum analyses were discarded. However, a minimum of two replicates per rum sample were always considered.

In this first step, the HCA study classified satisfactorily five samples containing additives such as honey, syrup or flavoring (labelled as vanilla and tangerine flavored rums). The dendrogram obtained is shown in Fig. 2. Their chromatographic profiles showed relevant differences regards to the rest of rum samples. Therefore, and after concluding about the easy discrimination of rum samples containing additives,

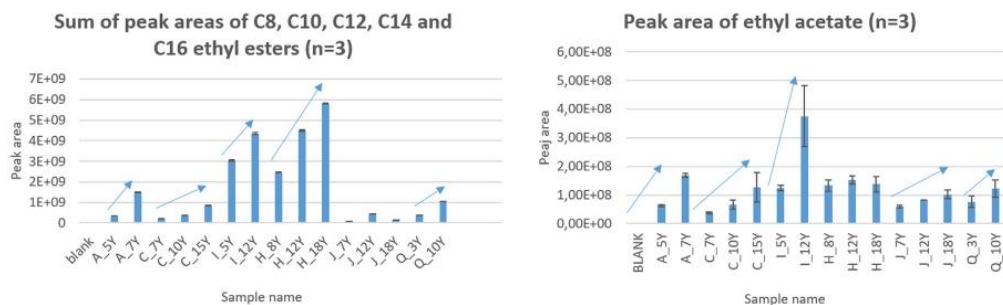


Fig. 1. Bar chart representation of ethyl acetate (left) and the sum ethyl esters between C8 and C16 (right) among different rum samples. Arrows were added to highlight aging trends within the same brands.

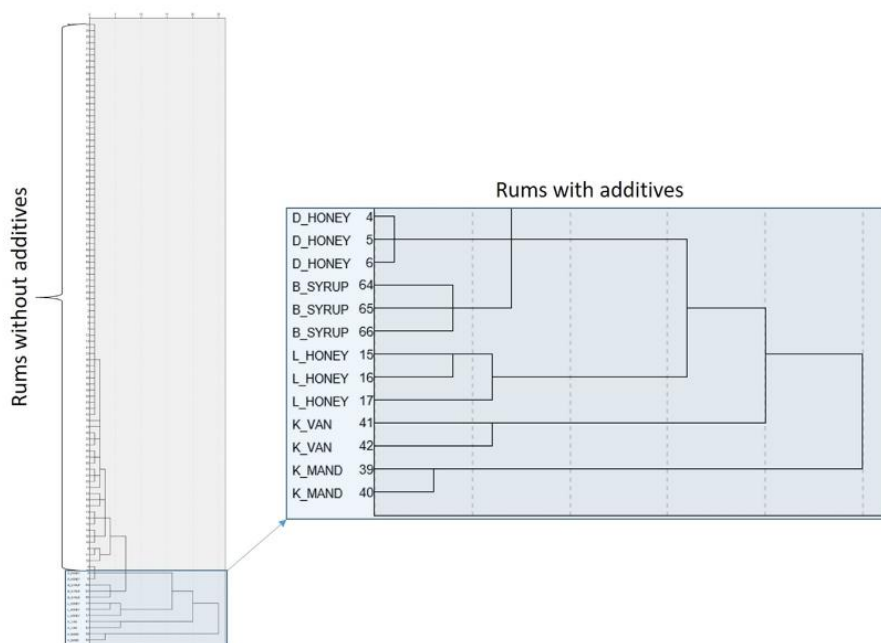


Fig. 2. Dendrogram obtained by HCA using the total rum batch. See Table 1 for sample code.

these kind of samples were disregarded to facilitate further chemometric classification of the rest of the rums.

Applying the factor analysis method, where the data was submitted to an orthogonal rotation (*varimax rotation*), a total of 40 variables (m/z) were selected as the most discriminant ions (score values higher than 0.80) for further PCA. The factor analysis scores obtained for all ions are shown in Table S-2 (Supporting Information).

PCA using the selected variables was applied to transform the high-dimensional variables into a small number of orthogonal factors, principal components (PCs), whilst accounting for the largest variance. PCA three-dimensional representation (Fig. 3) provided 88% of the total variance explained on the three first PCs. In this plot, PC1, PC2 and PC3 account for 36%, 36% and 15% of the individual variance, respectively.

3.5. Identification of the most discriminant compounds

These most discriminant ions were initially assigned to 13 volatile organic compounds based on their retention indexes and mass spectrometric data using the NIST database (match factor higher than 750). Further identification studies by HRMS with Q-Exactive Orbitrap were carried out and their mass spectra evaluated by a high-resolution filtering (HRF) tool from the Tracefinder software. Initially, the HRF assigned tentative identification was based on the use of high-resolution mass spectra but using the traditional spectral matching at unit resolution. Later, all unique combinations of atoms from these candidate precursors are generated and matched to m/z peaks using narrow mass tolerances [48]. The HRF scores obtained for the candidates ranged from 99.50% to 100.00%, confirming 11 of the compounds previously assigned by GC-QqQ-MS analysis using retention index and NIST database search (Table 2). Although different identification methods were applied for the same compound, further discussion was necessary for final confirmation of the three hesitant results.

The initial confirmation of ethyl acetate presented certain doubts because of the monitored mass range (m/z 50–400). Its predominant m/z 43 (CH_3CO^+) was not observed in those experimental conditions.

Therefore, a new m/z range was investigated (m/z 35–400) for this particular case obtaining an adequate library spectrum comparison (Fig. S3 in Supporting Information) and a 100% HRF score. However, it should be noted that the use of a lower mass range has caused significant interferences for high volatility compounds (retention time lower than 5 min) due to the presence of high concentrations of ethanol (Fig. S4 in Supporting Information).

Two of the studied chromatographic peaks (retention time 18.4 and 19.5 min) were initially identified by NIST as ionene (1,2,3,4-tetrahydro-1,1,6-trimethyl-naphthalene). However, HRF study of the HRMS data provided more reliable confirmation for the peak at 18.4 min (HRF score of 100.00%). The chromatographic peak at retention 19.5 min also presented a high HRF score (99.53%) and probably it is a ionene-derivative, but not included in the NIST library. Ionene has been reported as a pyrolysis degradation of carotenoid products [49] that have been reported as present in sugarcane [50]. Therefore, ionene and its related compound can be produced during molasses production.

For the identification of the compound at RT 11.8min, further discussion was required. After GC-QqQ-MS analysis, NIST library search showed trans-2-tetrahydro-5-methyl-furanmethanol as the main tentative identification option. However, further HRMS Q-Exactive Orbitrap analysis identified tetrahydro-2H-pyran-2-methanol as the first choice. Both compounds present the same formula ($\text{C}_6\text{H}_{12}\text{O}_2$), exact mass (m/z 116.08318) and very similar mass spectrometric profile. The main difference between both spectra is the distribution of ions in the cluster (m/z 55–60) shown in Fig. 4. The cluster profile obtained in the experimental spectrum fits with the one shown for tetrahydro-2H-pyran-2-methanol. This compound also presented a high HRF score of 100%.

Table 2 shows the 13 selected compounds with their corresponding identified compounds, empirical formula, selected ions, retention time, KI value, HRF score and identification methods used. It can be readily observed that all ions from PC1 and PC2 belong to six ethyl ester compounds (hexadecanoic acid, ethyl ester; (E)-9-octadecenoic acid, ethyl ester; tetradecanoic acid, ethyl ester; octanoic acid, 3-methylbutyl ester; decanoic acid, ethyl ester; octanoic acid, ethyl ester). Ethyl ester compounds, formed by the reaction of ethanol with acyl-CoA, are well-

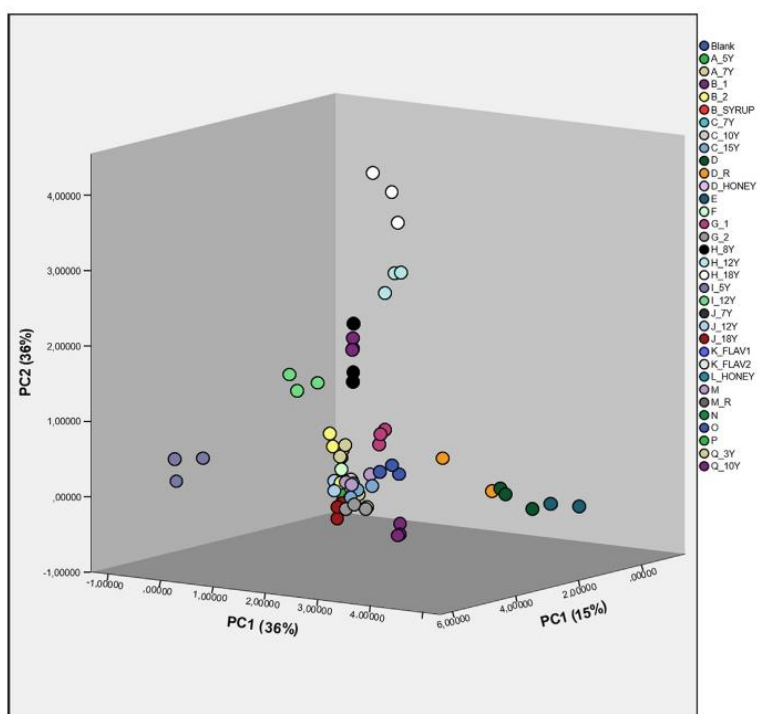


Fig. 3. PCA representation using the most discriminant ions.

known for playing an important role in the organoleptic properties of fermented beverages, due to their fruity aroma [51]. For PC3, high discrimination power was observed for low molecular weight compounds, such as ethyl acetate, diethoxymethane and 1,1 diethoxybutane. Ethyl acetate provides fruity and brandy notes to beverages [52]. Frantiza et al. has already reported the presence of 1,1-diethoxy-3-methyl-butane in two different rums, providing intense fruity aroma

[53]. Additionally, ionene and tetrahydro-2H-Pyran-2-methanol were also identified as highly discriminant in PC3.

3.6. Supervised analyses

After variables selection, LDA was applied in order to classify the rums according to different groups by maximizing the ratio of between-

Table 2
Most discriminant ions for PC1, PC2, and PC3 with their score, retention time and corresponding tentatively identified compounds.

PC	Compound	Empirical formula	<i>m/z</i> (score)	RT (min)	KI ^a	HRF (%) ^b	Identification method
1	Hexadecanoic acid, ethyl ester	C ₁₈ H ₃₆ O ₂	185.2 (0.94); 241.2 (0.94); 284.3 (0.93); 213.2 (0.93); 157.2 (0.89); 115.1 (0.88); 88.1 (0.86); 55.1 (0.85);	24.10	1993	100.00	KI. Lib. ^c , HRF
1	(E)-9-Octadecenoic acid ethyl ester	C ₂₀ H ₃₈ O ₂	180.2 (0.91); 112.2 (0.90); 222.3 (0.87); 264.2 (0.86); 137.1 (0.85); 109.1 (0.85); 169.1 (0.84)	25.17	2170	100.00	KI. Lib., HRF
1	Tetradecanoic acid, ethyl ester	C ₁₆ H ₃₂ O ₂	88.1 (0.86); 55.1 (0.82)	23.05	1789	100.00	KI. Lib., HRF
2	Octanoic acid, 3-methylbutyl ester	C ₁₃ H ₂₆ O ₂	70.2 (0.95); 127.2 (0.92)	20.89	1441	99.83	KI. Lib., HRF
2	Decanoic acid, ethyl ester	C ₁₂ H ₂₄ O ₂	157.1 (0.92); 115.1 (0.92); 88.1 (0.81); 55.1 (0.91)	20.44	1389	99.50	KI. Lib., HRF
2	Octanoic acid, ethyl ester	C ₁₀ H ₂₀ O ₂	88.1 (0.85); 127.2 (0.85); 57.2 (0.85)	18.17	1194	99.86	KI. Lib., HRF
3	Ionene	C ₁₃ H ₁₈	131.1 (0.94); 159.1 (0.91); 144.1 (0.90); 116.1 (0.85); 91.1 (0.85); 113.1 (0.83)	18.4	1211	100.00	KI. Lib., HRF
3	Butane, 1,1-diethoxy-3-methyl-	C ₉ H ₂₀ O ₂	103.1 (0.94); 73.1 (0.94); 75.1 (0.94)	9.2	918	99.82	KI. HRF
3	Ionene-derivative	C ₁₃ H ₁₈	159.1 (0.90)	19.5	1295	99.53	HRF
3	Unknown	-	130.2 (0.89)	18.2	1197	-	-
3	Ethyl Acetate	C ₄ H ₈ O ₂	54.7 (0.87)	2.3	611	100.00	KI. Lib., HRF
3	2H-Pyran-2-methanol, tetrahydro-	C ₆ H ₁₂ O ₂	85.1 (0.86)	11.8	988	100.00	KI. Lib., HRF
3	Methane, 1,1-diethoxy-	C ₅ H ₁₂ O ₂	103.1 (0.84)	3.3	668	-	KI. Lib

^a KI; Kovats retention indices.

^b HRF (High-Resolution Filtering score); percentage of the spectrum obtained by HRMS Orbitrap that can be explained by combination of accurate mass, library matching and percentage of explained ions observed.

^c Lib; Identification based on mass spectrometric data using the NIST database (match factor higher than 750).

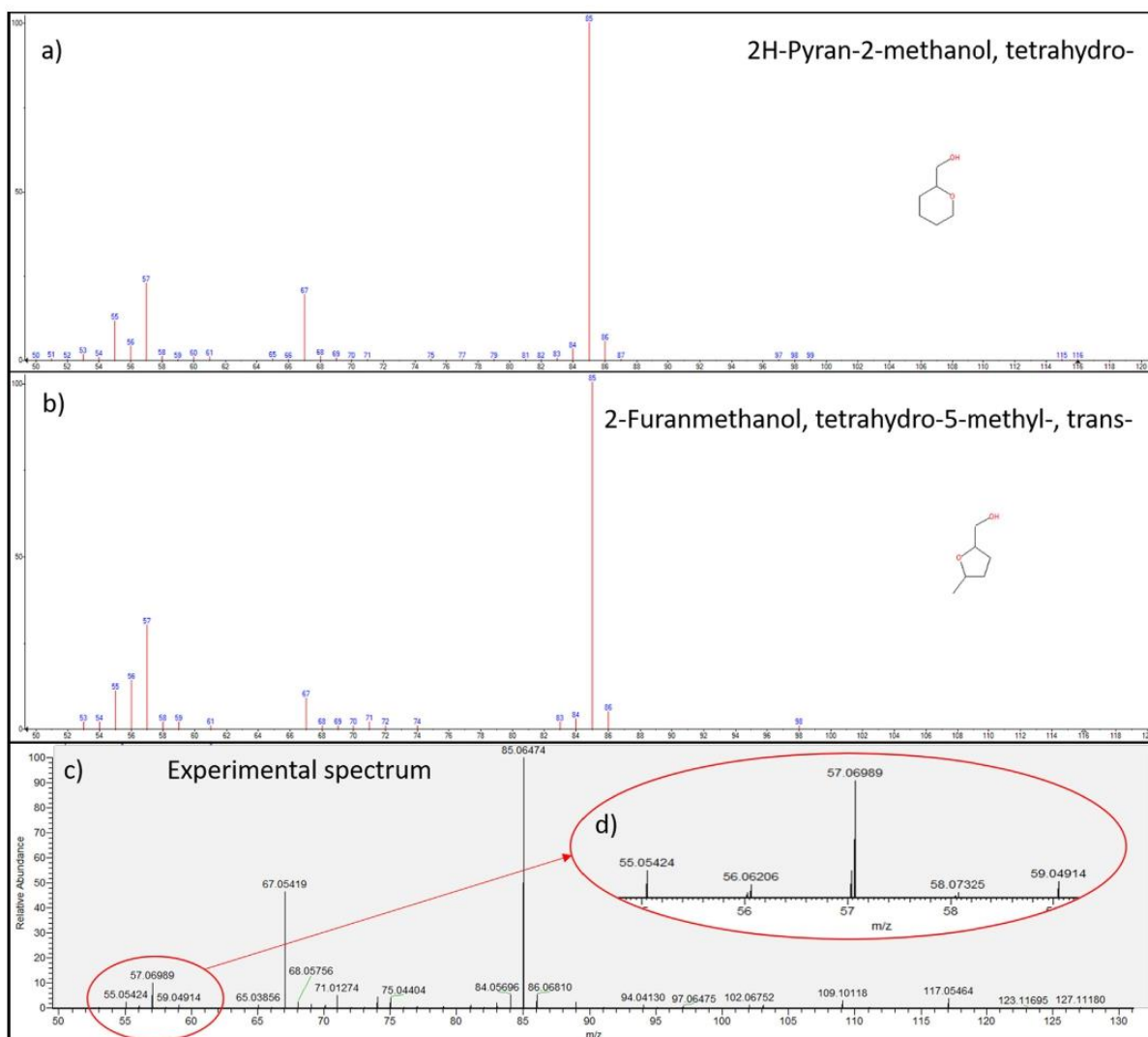


Fig. 4. Comparison between theoretical mass spectrum of tetrahydro-2H-pyran-2-methanol (a), theoretical mass spectrum of trans-tetrahydro-5-methyl-2-furanmethanol (b) and the obtained experimental spectrum by HRMS Q-Exactive Orbitrap (c) with zoom-in on the m/z 55–60 cluster (d).

class variance and minimizing the ratio of within-class variance. Classification groups were selected based on the main information about the rum elaboration, typically provided by rum manufacturers.

The leave-one-out method [54] was used as cross-validation procedure to evaluate the prediction ability (Q^2) for each LDA model using the previously selected variables. For a good predictability, the difference between R^2 (coefficient of determination) and Q^2 value should not exceed 0.3 and poor robustness of the model is usually suspected when that happens [55,56]. The rums that had no production information available for the respective classifying group were not used to ensure an accurate prediction.

For the country group, the 94% of the rums with 10 different origins were correctly classified ($R^2 = 0.94$) (Fig. S-5 in Supporting Information). When cross-validation was carried out, the LDA model had a prediction ability of $Q^2 = 0.73$. This cross-validation value can be explained because the composition of rum is mostly dependent on the manufacturing process due to its complexity and variability, rather than

the manufacturing country.

When raw material was used as classificatory criterion (Fig. 5.a), 97% of rums were correctly classified. The molasses group was clearly separated from the sugar cane group, while the group containing the rums where no information was available were placed close to the molasses group. The LDA data distribution reveals that the NA rums with no information about their production process could be elaborated from molasses. The cross-validation process, which used only the sugar cane group and the molasses group, provided a prediction ability $Q^2 = 0.91$ indicating that clear differences between these two groups provide a good prediction ability of this model.

When the classification was performed according to distillation method, where four categories were used (copper pot stills, stainless steel columns, a mixture, and NA) (Fig. S-5 in Supporting Information), 92% of the rums were correctly classified. The cross-validation showed a prediction ability of $Q^2 = 0.87$ of the rums. Rums made in copper pot stills were clearly differentiated from the rest. The other three

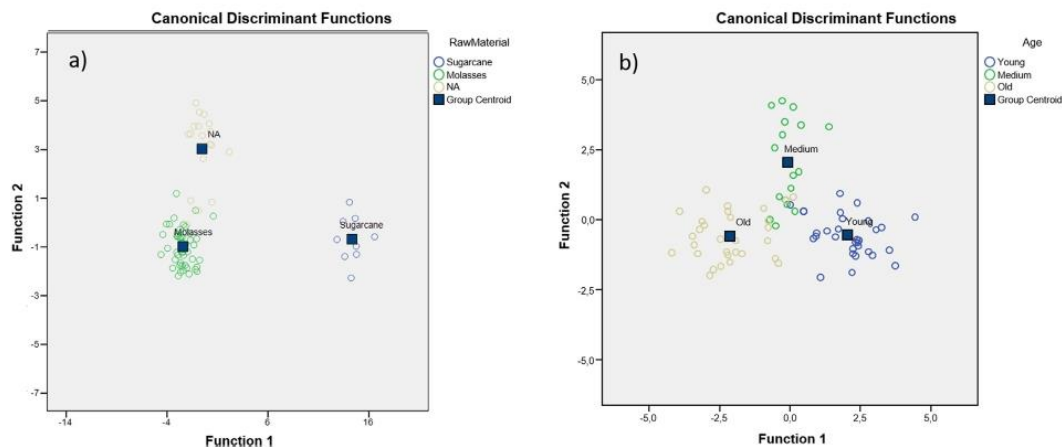


Fig. 5. LDA representation showing the classification of 28 rums based on (a) raw material and (b) aging period.

categories were correctly classified in most of the cases, despite their proximity in the LDA representation. Copper pot stills are the traditional method to distil rums; however, these days, many manufactures prefer stainless steel columns for their increased efficiency. With steel columns, only one fraction is collected, whereas copper pot stills provide manufacturers with the possibility to combine the head, heart and tail fractions to obtain specific rums. This elaboration difference can explain the clear LDA distinction of the copper pot distillation data.

For barrel type (unspecified oak barrel, American oak barrel, French oak barrel, NA), 95% of rums were correctly classified rums (Fig. S-5 in Supporting Information). A cross-validation value of $Q^2 = 0.86$ was obtained. The graphical proximity of the oak and American oak barrels could be an indication that American oak is the typical wood used when the manufacturers do not specify the type of barrel utilized.

The last LDA representation (Fig. 5.b) was based on rum aging (young, medium, old). The rum aging parameter was created by considering rums aged less than 5 years as young, those aged between 5 and 10 years as medium, and rums aged for more than years as old. A 94% of correctly classified rums were obtained for aging. For cross-validation, $Q^2 = 0.65$ were obtained. This value can be explained by the blending process and the differences in legislation about age labelling.

4. Conclusions

The present study has developed an analytical strategy for the classification of rums depending on different steps from the production process using the most discriminant compounds of the volatile fraction. For that, the targeted and untargeted analysis of rums was evaluated using HS-SPME-GC-MS and chemometric tools. The target analysis found some chemical indicators (ethyl acetate and ethyl esters of carboxylic acids) that could be correlated with aging within the same brand, but presented clear limitations when they were used across different brands. No other correlations with other parameters were found. On the other hand, the untargeted analysis using chemometric tools led to the classification of the 33 rums of different brands and ages. For that, unsupervised (HCA, PCA) and supervised techniques (LDA) were employed. The HCA showed considerable differences between the traditional rums and the rums prepared by addition of honey, syrup and flavoring. This chemometric method offers the potential to clearly distinguish these rums with additives from the rest. For the correct classification of traditional rums, PCA provided 40 ions as relevant chemical descriptors corresponding to 13 discriminant compounds (e.g. hexadecanoic acid ethyl ester, octanoic acid ethyl ester, decanoic acid ethyl ester, ethyl acetate and 1,1-diethoxy-3-methyl-

butane). For the confirmation of the compounds, a strategy based on the combination of retention indexes, NIST database matching using low-resolution mass spectrometry and HRF scores using high-resolution spectra obtained by HRMS Q-Exactive Orbitrap was employed.

The 28 traditional rums were classified based on manufacturing country, raw material, distillation method, barrel type, and aging period with data classification values from $R^2 = 0.94$ to $R^2 = 0.97$ (prediction ability from $Q^2 = 0.65$ to $Q^2 = 0.91$). LDA results showed conclusive differences between the classification groups and the importance of enough representative sample for each group. The applied classification strategy allowed for a better understanding of rum composition, including for rums with not accurate label information. Moreover, this technique could be of interest to future investigations, for example, to other spirit beverages.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.talanta.2018.05.025>.

References

- [1] Authentic Caribbean Rum, Decade of rum, <<https://www.acr-rum.com/assets/pdf/decade-of-rum-english-usa.pdf>>, 2013.
- [2] Top 25 Countries by Rum Consumption Per Capita, (n.d.). <<https://blog.euromonitor.com/2017/08/top-25-countries-by-rum-consumption-per-capita.html>> (Accessed 20 March 2018).
- [3] L. Frantiza, M. Granvogl, P. Schieberle, Influence of the production process on the key aroma compounds of rum: from molasses to the spirit, *J. Agric. Food Chem.* 64 (2016) 9041–9053, <http://dx.doi.org/10.1021/acs.jafc.6b04046>.
- [4] J.A. Pino, S. Tolle, R. Gök, P. Winterhalter, Characterisation of odour-active compounds in aged rum, *Food Chem.* 132 (2012) 1436–1441, <http://dx.doi.org/10.1016/j.foodchem.2011.11.133>.
- [5] D.R. Cardoso, L.G. Andrade-Sobrinho, A.F. Leite-Neto, R.V. Reche, W.D. Isique, M.M.C. Ferreira, B.S. Lima-Neto, D.W. Franco, Comparison between cachaça and rum using pattern recognition methods, *J. Agric. Food Chem.* 52 (2004) 3429–3433, <http://dx.doi.org/10.1021/jf035262+>.
- [6] J. Pino, M.P. Martí, M. Mestres, J. Pérez, O. Busto, J. Guasch, Headspace solid-phase microextraction of higher fatty acid ethyl esters in white rum aroma, *J. Chromatogr. A* 954 (2002) 51–57, [http://dx.doi.org/10.1016/S0021-9673\(02\)00167-X](http://dx.doi.org/10.1016/S0021-9673(02)00167-X).
- [7] D.A. Nicol, Rum, in: A.G.H. Lea, J. Piggott (Eds.), *Fermented Beverage Prod.* Springer, US, 2003, pp. 263–287.
- [8] J.E. Murtagh, Feedstocks, fermentation and distillation for production of heavy and light rums, *alcohol Textb. A Ref. beverage, Fuel Ind. Alcohol Ind.* (2007) 243–255.
- [9] P. Wiśniowska, M. Śliwińska, T. Dymerski, W. Wardencki, J. Namieśnik, Application of gas chromatography to analysis of spirit-based alcoholic beverages, *Crit. Rev.*

- Anal. Chem. 45 (2015) 201–225, <http://dx.doi.org/10.1080/10408347.2014.904732>.
- [10] D.W. Lachenmeier, R. Attig, W. Frank, C. Athanasakis, The use of ion chromatography to detect adulteration of vodka and rum, *Eur. Food Res. Technol.* 218 (2003) 105–110, <http://dx.doi.org/10.1007/s00217-003-0799-8>.
 - [11] P. Hernández-Orte, E. Franco, C.G. Huerta, J.M. García, M. Cabellos, J. Suberviola, I. Orriols, J. Cacho, Criteria to discriminate between wines aged in oak barrels and macerated with oak fragments, *Food Res. Int.* 57 (2014) 234–241, <http://dx.doi.org/10.1016/j.foodres.2014.01.044>.
 - [12] CFR, - Code of Federal Regulations of the United States of America, Alcohol, Tobacco Products and Firearms, Washington DC, 1991.
 - [13] European Parliament Council of the European Union, 110/2008 of the European parliament and of the council of 15 January 2008 on the definition, description, presentation, labelling and the protection of geographical indications of spirit drinks and repealing Council Regulation (EEC) No 1576/89 Off. 2008.
 - [14] Canada Department of Justice, Food and Drug Regulations C.R.C. c 870, 2009.
 - [15] A.E. Springer, J. Riedl, S. Esslinger, T. Roth, M.A. Glomb, C. Fauth-Hasek, Validated modeling for german white wine varietal authentication based on headspace solid-phase microextraction online coupled with gas chromatography mass spectrometry fingerprinting, *J. Agric. Food Chem.* 62 (2014) 6844–6851, <http://dx.doi.org/10.1021/jf502042c>.
 - [16] Z. Xiao, S. Liu, Y. Gu, N. Xu, Y. Shang, J. Zhu, Discrimination of cherry wines based on their sensory properties and aromatic fingerprinting using HS-SPME-GC-MS and multivariate analysis, *J. Food Sci.* 79 (2014) 284–294.
 - [17] Y. Niu, X. Zhang, Z. Xiao, S. Song, K. Eric, C. Jia, H. Yu, J. Zhu, Characterization of odor-active compounds of various cherry wines by gas chromatography–mass spectrometry, gas chromatography–olfactometry and their correlation with sensory attributes, *J. Chromatogr. B* 879 (2011) 2287–2293.
 - [18] S. Grützmann, V. Callari, M. Sganzerla, H. Teixeira, Talanta Volatile composition of Merlot red wine and its contribution to the aroma: optimization and validation of analytical method, *Talanta* 174 (2017) 752–766, <http://dx.doi.org/10.1016/j.talanta.2017.06.074>.
 - [19] A.K. Hjelmeland, P.L. Wylie, S.E. Ebeler, Talanta A comparison of sorptive extraction techniques coupled to a new quantitative, sensitive, high throughput GC – MS / MS method for methoxypyrazine analysis in wine, *Talanta* 148 (2016) 336–345, <http://dx.doi.org/10.1016/j.talanta.2015.10.086>.
 - [20] J. Jiao, N. Ding, T. Shi, X. Chai, P. Cong, Z. Zhu, Study of Chromatographic Fingerprint of the Flavor in Beer by HS-SPME-GC, *Anal. Lett.* 44 (2011) 648–655.
 - [21] O. Pinho, I.M. Ferreira, L.H. Santos, Method optimization by solid-phase microextraction in combination with gas chromatography with mass spectrometry for analysis of beer volatile fraction, *J. Chromatogr. A* 1121 (2006) 145–153.
 - [22] F. Rodrigues, M. Caldeira, J. de S. Câmara, Development of a dynamic headspace solid-phase microextraction procedure coupled to GC–qMSD for evaluation the chemical profile in alcoholic beverages, *Anal. Chim. Acta* 609 (2008) 82–104.
 - [23] D. Saison, D.P. De Schutter, F. Delvaux, F.R. Delvaux, Optimisation of a complete method for the analysis of volatiles involved in the flavour stability of beer by solid-phase microextraction in combination with gas chromatography and mass spectrometry, *J. Chromatogr. A* 1190 (2008) 342–349.
 - [24] N. Moreira, S. Meireles, T. Brandão, P. Guedes, D. Pinho, Talanta Optimization of the HS-SPME – GC – IT / MS method using a central composite design for volatile carbonyl compounds determination in beers, *Talanta* 117 (2013) 523–531, <http://dx.doi.org/10.1016/j.talanta.2013.09.027>.
 - [25] L. Rong, L.J. Peng, C.T. Ho, S.H. Yan, M. Meurens, Z.Z. Zhang, D.X. Li, X.C. Wan, G.H. Bao, X.L. Gao, T.J. Ling, Brewing and volatiles analysis of three tea beers indicate a potential interaction between tea components and lager yeast, *Food Chem.* 197 (2016) 161–167, <http://dx.doi.org/10.1016/j.foodchem.2015.10.088>.
 - [26] C.N. Rhodes, K. Heaton, I. Goodall, P.A. Brereton, Gas chromatography carbon isotope ratio mass spectrometry applied to the detection of neutral alcohol in Scotch whisky: an internal reference approach, *Food Chem.* 114 (2009) 697–701.
 - [27] J.I. Cacho, N. Campillo, P. Viñas, M. Hernández-córdoba, Talanta cloud point extraction and gas chromatography with direct microvial insert thermal desorption for the determination of haloanisoles in alcoholic beverages, *Talanta* 160 (2016) 282–288, <http://dx.doi.org/10.1016/j.talanta.2016.07.029>.
 - [28] S. Vichi, M. Riu-Aumatell, M. Mora-Pons, J.M. Guadayol, S. Buxaderas, E. López-Tamames, HS-SPME coupled to GC/MS for quality control of Juniperus communis L. berries used for gin aromatization, *Food Chem.* 105 (2007) 1748–1754.
 - [29] A.J. Johnson, H. Heymann, S.E. Ebeler, Volatile and sensory profiling of cocktail bitters, *Food Chem.* 179 (2015) 343–354, <http://dx.doi.org/10.1016/j.foodchem.2015.01.114>.
 - [30] E.S.P. Nascimento, D.R. Cardoso, D.W. Franco, Quantitative ester analysis in cacha (ç) and distilled spirits by gas chromatography – mass spectrometry (GC – MS), *J. Agric. Food Chem.* 56 (2008) 5488–5493.
 - [31] O.M. Sampaio, R.V. Reche, D.W. Franco, Chemical profile of rums as a function of their origin. The use of chemometric techniques for their identification, *J. Agric. Food Chem.* 56 (2008) 1661–1668, <http://dx.doi.org/10.1021/jf0726841>.
 - [32] F.W.B. Aquino, L.M. Boso, D.R. Cardoso, D.W. Franco, Amino acids profile of sugar cane spirit (cacha(ç)a), rum, and whisky, *Food Chem.* 108 (2008) 784–793.
 - [33] Z.L. Cardeal, P.J. Marriott, Comprehensive two-dimensional gas chromatography-mass spectrometry analysis and comparison of volatile organic compounds in Brazilian cacha(ç)a and selected spirits, *Food Chem.* 112 (2009) 747–755, <http://dx.doi.org/10.1016/j.foodchem.2008.06.057>.
 - [34] E.A. Nonato, F. Carazza, F.C. Silva, C.R. Carvalho, Z. de L. Cardeal, A headspace solid-phase microextraction method for the determination of some secondary compounds of Brazilian sugar cane spirits by gas chromatography, *J. Agric. Food Chem.* 49 (2001) 3533–3539.
 - [35] J.A. Pino, Characterization of rum using solid-phase microextraction with gas chromatography-mass spectrometry, *Food Chem.* 104 (2007) 421–428, <http://dx.doi.org/10.1016/j.foodchem.2006.09.031>.
 - [36] R.S. da Costa, S.R.B. Santos, L.F. Almeida, E.C.L. Nascimento, M.J.C. Pontes, R.A.C. Lima, S.S. Simões, M.C.U. Araújo, A novel strategy to verification of adulteration in alcoholic beverages based on Schlieren effect measurements and chemometric techniques, *Microchem. J.* 78 (2004) 27–33.
 - [37] J. Bouhleb, D.J. Bouveresse, S. Abouelkaram, E. Baéza, C. Jondreville, A. Travel, J. Ratel, E. Engel, D.N. Rutledge, Talanta comparison of common components analysis with principal components analysis and independent components analysis: application to SPME-GC-MS volatolomic signatures, *Talanta* 178 (2018) 854–863, <http://dx.doi.org/10.1016/j.talanta.2017.10.025>.
 - [38] U. Römisch, H. Jäger, X. Capron, S. Lanteri, M. Forina, J. Smeyers-Verbeke, Characterization and determination of the geographical origin of wines. Part III: multivariate discrimination and classification methods, *Eur. Food Res. Technol.* 230 (2009) 31–45.
 - [39] A. Montaña, Talanta volatile profiles of green Spanish-style table olives: application of compositional data analysis for the segregation of their cultivars and production areas, *Talanta* 169 (2017) 77–84, <http://dx.doi.org/10.1016/j.talanta.2017.03.066>.
 - [40] A. Biancolillo, R. Bucci, A.L. Magri, A.D. Magri, F. Marini, Data-fusion for multi-platform characterization of an Italian craft beer aimed at its authentication, *Anal. Chim. Acta* 820 (2014) 23–31.
 - [41] J. Saurina, Characterization of wines using compositional profiles and chemometrics, *TrAC Trends Anal. Chem.* 29 (2010) 234–245.
 - [42] F. Magagna, L. Valverde-Som, C. Ruiz-Sambal, A.S. L. Cuadros-Rodríguez, S.E. Reichenbach, C. Bicchi, C. Cordero, Combined untargeted and targeted fingerprinting with comprehensive two-dimensional chromatography for volatiles and ripening indicators in olive oil, *Anal. Chim. Acta* (2016), <http://dx.doi.org/10.1016/j.aca.2016.07.005>.
 - [43] V.I. Babushok, Trends in analytical chemistry chromatographic retention indices in identification of chemical compounds, *Trends Anal. Chem.* 69 (2015) 98–104, <http://dx.doi.org/10.1016/j.trac.2015.04.001>.
 - [44] P. Wenig, J. Odermatt, OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data, *BMC Bioinform.* 11 (2010) 405, <http://dx.doi.org/10.1186/1471-2105-11-405>.
 - [45] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresić, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, *BMC Bioinform.* 11 (2010) 395, <http://dx.doi.org/10.1186/1471-2105-11-395>.
 - [46] S. Risticic, D. Vuckovic, J. Pawliszyn, Solid - Phase Microextraction development of fiber solid - phase, 2010.
 - [47] A. Ruiz-Delgado, F.J. Arrebola-Liébana, R. Romero-González, R. López-Ruiz, A. Garrido-Frenich, Headspace solid-phase microextraction coupled to gas chromatography-tandem mass spectrometry for the determination of haloanisoles in sparkling (cava and cider) and non-sparkling (wine) alcoholic beverages, *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* 33 (2016) 1535–1544, <http://dx.doi.org/10.1080/19440049.2016.1229870>.
 - [48] N.W. Kwiecien, D.J. Bailey, M.J.P. Rush, J.S. Cole, A. Ulbrich, A.S. Hebert, M.S. Westphall, J.J. Coon, High-resolution filtering for improved small molecule identification via GC/MS, *Anal. Chem.* 87 (2015) 8328–8335, <http://dx.doi.org/10.1021/acs.analchem.5b01503>.
 - [49] R.G. Achari, G. Shaw, R. Villarhead, Identification of ionene and other carotenoid degradation products from the pyrolysis of sporopollenins derived from some pollen exines, a spore coal and the Green River shale, *Chem. Geol.* 12 (1973) 229–234, [http://dx.doi.org/10.1016/0009-2541\(73\)90091-0](http://dx.doi.org/10.1016/0009-2541(73)90091-0).
 - [50] J.D. Fontana, B. Czezuga, T.M.B. Bonfim, M.B. Chociai, B.H. Oliveira, M.F. Guimarães, M. Baron, Bioproduction of carotenoids: the comparative use of raw sugarcane juice and depolymerized bagasse by *Phaffia rhodozyma*, *Bioresour. Technol.* 58 (1996) 121–125, [http://dx.doi.org/10.1016/S0960-8524\(96\)00092-2](http://dx.doi.org/10.1016/S0960-8524(96)00092-2).
 - [51] K.J. Verstrepen, G. Derdelinckx, J.P. Dufour, J. Winderickx, J.M. Thevelein, I.S. Pretorius, F.R. Delvaux, Flavor-active esters: adding fruitiness to beer, *J. Biosci. Bioeng.* 96 (2003) 110–118, [http://dx.doi.org/10.1016/S1389-1723\(03\)90112-5](http://dx.doi.org/10.1016/S1389-1723(03)90112-5).
 - [52] A.P. Dionísio, G. Molina, D. Souza de Carvalho, R. Dos Santos, J.L. Bicas, G.M. Pastore, Natural flavourings from biotechnology for foods and beverages, *Nat. Food Addit. Ingrid. Flavour.* (2012) 231–259, <http://dx.doi.org/10.1016/B978-1-84569-811-9.50011-6>.
 - [53] L. Franitz, M. Granvogel, P. Schieberle, Characterization of the key aroma compounds in two commercial rums by means of the sensomics approach, *J. Agric. Food Chem.* 64 (2016) 637–645, <http://dx.doi.org/10.1021/acs.jafc.5b05426>.
 - [54] D.W. Ostens, Selection of optimal regression models via cross-validation, *J. Chemom.* 2 (1988) 39–48, <http://dx.doi.org/10.1002/cem.1180020106>.
 - [55] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Ravichandran Veerasamy, et al., Validation of QSAR models-strategies and importance validation of qsar models -strategies and importance, (n.d.). <https://pdfs.semanticscholar.org/4eb2/5ff5a87f2fd6789c5b9954eddddfdc59dab.pdf> (Accessed 24 January 2018).
 - [56] B. Worley, R. Powers, HHS Public Access, 1 (2015) pp. 92–107. doi:10.2174/2213235X11301010092.Multivariate.




2.4 Aportación científica 2. An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography–orbitrap mass spectrometry and chemometric strategies

J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido Frenich. *J. Agric. Food Chem.* 67 (2019) 1302–1311.

DOI: 10.1021/acs.jafc.8b05622

CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON

An Innovative Metabolomic Approach for Golden Rum Classification Combining Ultrahigh-Performance Liquid Chromatography–Orbitrap Mass Spectrometry and Chemometric Strategies

José Raúl Belmonte-Sánchez, Roberto Romero-González, Francisco Javier Arrebola, José Luis Martínez Vidal, and Antonia Garrido Frenich*

Research Group “Analytical Chemistry of Contaminants”, Department of Chemistry and Physics, Research Centre for Agricultural and Food Biotechnology (BITAL), University of Almería, Agrifood Campus of International Excellence, ceiA3, E-04120 Almería, Spain

Supporting Information

ABSTRACT: A comprehensive fingerprinting strategy for golden rum classification considering different categories such as fermentation barrel, raw material, and aging is provided, using a metabolomic fingerprinting approach. A nontarget fingerprinting of 30 different rums using liquid chromatography coupled to high-resolution mass spectrometry (Exactive Orbitrap mass analyzer, LC-HRMS) was applied. Principal component analysis (PCA) was used to assess the overall structure of the data and to identify potential outliers. Different chemometric analyses such as partial least-squares discriminant analysis (PLS-DA) were used. A variable importance in projection (VIP) selection method was applied to identify the most significant markers that allow group separation. Compounds related to aging and fermentation processes such as furfural derivatives (e.g., hydroxymethylfurfural) and sugars (e.g., glucose, mannitol) were found as the most discriminant compounds (VIP threshold value >1.5). Suitable separation according to selected categories was achieved, and a classification ability of the models of close to 100% was achieved.

KEYWORDS: rum, metabolomics, liquid chromatography, high-resolution mass spectrometry, fingerprinting, multivariate analysis

INTRODUCTION

Rum is one of the most popular spirit-based drinks, taking into account a high world consumption rate (more than 1 billion of liters per year) and an expected increase of 1.9% in volume terms over 2016–2021. This alcoholic drink is exclusively obtained from the alcoholic fermentation and distillation of either molasses or sugar cane syrups. Golden rums are the most popular, and their color is due to oak maturation. There is no formal designation of what makes a “golden” rum, but these are generally mellow.¹

Rum’s production begins with the selection of the raw material. Once the raw material is selected, sugar cane juice or molasses is placed in large vats, and ethanol is produced from the natural sugars by the action of yeast enzymes. The resulting mash is distilled using heat in copper pot stills or in stainless steel columns to obtain a high content of ethanol. The principal compounds obtained by distillation are water, ethanol, and hundreds of organic characteristic compounds, which contribute to distillate flavor.^{2,3} Then, pure demineralized water is used for dilution of the resulting distillate to obtain a desired alcohol percentage of 37.5–40.0% (minimum 37.5% v/v imposed by EU regulation for rum beverages).⁴ The resulting solution is then aged in oak barrels previously used for whiskey, wine, or brandy production.⁵ American white oak (*Quercus alba*) and French red oak (*Q. robur* and *Q. petraea*) are the most frequently used sources of oak woods,⁶ although the basic composition of oak does not significantly differ from one species to another. Blending, which gives to the rum particular sensory characteristics in an optional step, can also

be applied in rum production. Rums of different ages are mixed to obtain specific organoleptic products. All these types of processes make rum an attractive object of study.

Currently, different methods have been described for the classification of alcoholic beverages on the basis of an analysis of the metabolomic composition,⁷ including scarce rum classification studies.^{8,9} For that purpose, gas chromatography coupled to mass spectrometry (GC-MS) has been one of the most frequently used techniques for analyzing the volatile fraction of the rums.

However, rum studies applying liquid chromatography coupled to mass spectrometry (LC-MS) have been less frequent, and the nonvolatile fractions of this beverage have been mainly analyzed. Other rum studies have been generally limited to the identification of a few compounds such as phenolic compounds, formaldehyde, and monoterpenes.^{10–12} Several alcoholic matrices such as whiskey,^{13–15} brandy,¹⁶ vodka,^{17,18} beer,^{19,20} and wine^{21,22} have been commonly studied in terms of authentication, quality control, differentiation, and adulteration detection, generally using LC-MS.²³ Due to the complexity and variability of rum preparation, their classification represents an analytical challenge.

With this aim, chemometric methods coupled with analytical techniques are being increasingly used for beverage studies.²³

Received: October 12, 2018

Revised: January 4, 2019

Accepted: January 8, 2019

Published: January 8, 2019

Table 1. Code, Bottle Labeling, and Manufacturer Information for the Rum Samples Analyzed

brand code	bottle code	origin	aging	raw material	barrel
TAB	TAB	Caribbean	young	N/A	N/A
HAV	HAV_5	Cuba	5 years	molasses	French oak barrel
	HAV_7		7 years	molasses	French oak barrel
LEG	LEG_A		anejo	molasses	American oak barrel
	LEG_B		young	molasses	no barrel
MAT	MAT_7	Dominican Republic	7 years	molasses	American oak barrel
	MAT_10		10 years	molasses	American oak barrel
	MAT_15		15 years	molasses	American oak barrel
ALM	ALM_R1		anejo	N/A	American oak barrel
	ALM_R2		anejo	N/A	American oak barrel
	ALM_B		young	N/A	no barrel
BAR	BAR		old reserve	N/A	N/A
BOH	BOH		anejo	N/A	N/A
PLA	PLA_G	Grenada	double aged, old reserve	molasses	mixed (American + French oak barrel)
	PLA_T	Trinidad & Tobago	double aged, overproof	molasses	mixed (American + French oak barrel)
BOT	BOT_8	Guatemala	8 years	sugar cane juice	American oak barrel
	BOT_12		12 years	sugar cane juice	American oak barrel
	BOT_18		18 years	sugar cane juice	American oak barrel
ZAC	ZAC		6–25 years	sugar cane juice	American oak barrel
APP	APP_5	Jamaica	5–10 years	molasses	American oak barrel
	APP_12		12 years	molasses	American oak barrel
FLOR	FLOR_7	Nicaragua	7 years	molasses	American oak barrel
	FLOR_12		12 years	molasses	American oak barrel
	FLOR_18		18 years	molasses	American oak barrel
NEG	NEG_R1	Spain	anejo	molasses	N/A
	NEG_R2		molasses	N/A	
REY	REY		N/A	N/A	no barrel
ST	ST_3	Venezuela	3 years	molasses	French oak barrel
	ST_10		10 years	molasses	French oak barrel
VEL	VEL		young	N/A	N/A

Different and very diverse supervised and unsupervised techniques such as principal component analysis (PCA), hierarchical clustering analysis (HCA), liner discriminant analysis (DA), partial least-squares discriminant analysis (PLS-DA), and orthogonal partial least-squares discriminant analysis (OPLS-DA) have been already used for the chemometric evaluation of food and beverage quality.^{24–26} Unsupervised methods (e.g., PCA and HCA) have usually been used in exploring the overall structure of a data set, finding trends, and grouping within the data set.^{27,28} Supervised techniques are applied to support a priori known data structures and to train patterns and rules to predict new data, which can be classified using linear methods (e.g., PLS-DA, LDA, OPLS-DA).^{29,30}

Herein, this study presents, for the first time, a comprehensive fingerprinting strategy for golden rum classification on the basis of different categories such as barrel, raw material, and aging. With this aim, a nontarget fingerprint approach based on LC coupled to high-resolution MS (Exactive-Orbitrap mass analyzer, LC-HRMS) was applied for the classification of 30 commercially available rums. PCA models were built to assess the overall structure of the metabolite data and to identify potential outliers. Several chemometric analyses such as PLS-DA were used to build discriminant models. A variable importance in projection (VIP) method was applied to identify the most significant markers that allow group separation. For the identification of markers, metabolite annotation was determined from the exact mass composition, from the goodness of isotopic fit of the

predicted molecular formula and matching with available databases.

MATERIALS AND METHODS

Reagents. Anhydrous ammonium acetate (>98%) was purchased from Panreac (Barcelona, Spain). Ultrapure water was obtained by a Milli-Q water gradient system (Millipore, Bedford, MA, USA). Formic acid was obtained from Fisher Scientific (Erembodegem, Belgium), and methanol (MeOH) (LC-MS grade) was acquired from Fluka (St. Louis, MO, USA).

Samples. For this study, 30 commercial rums were purchased from different local liquor stores located in Almeria, Spain. The rums were manufactured in 10 different countries: Cuba (4 samples), Dominican Republic (8 samples), Grenada (1 sample), Guatemala (4 samples), Jamaica (2 samples), Nicaragua (3 samples), Caribbean region (1 sample), Spain (3 samples), Trinidad & Tobago (1 sample), and Venezuela (3 samples). All samples were stored at room temperature prior to analysis, in their original glass bottles. Information about the rum production, obtained from the official Web site of rum manufacturers as well as from the label, and assigned codes for each rum are summarized in Table 1. It should be pointed out that information about aging, raw material, and barrel was not provided by all manufacturers. When the information was not available, this was recorded as N/A.

Sample Preparation Procedure. A simple pretreatment method was carried out prior to LC-HRMS analysis. Thus, 1 mL of rum sample was diluted up to 5 mL with Milli-Q water and shaken for 1 min in a vortex. Then, 10 μ L was directly injected into the analyzer.

Rums from the same brand were analyzed equally across the sampling sequence according to a block design in order to guarantee their comparability and lack of potential analytical bias. Three replicates of each bottle were analyzed.

LC-HRMS Analysis. For the UHPLC-Orbitrap-MS analysis, a Thermo Fisher Scientific Transcend 600 LC instrument (Thermo Scientific Transcend, Thermo Fisher Scientific, San Jose, CA, USA) was used and it was coupled to a Orbitrap Thermo Fisher Scientific mass spectrometer (Exactive, Thermo Fisher Scientific, Bremen, Germany) using an electrospray interface (ESI) (HESI-II, Thermo Fisher Scientific, San Jose, CA, USA). The chromatographic separation was carried out with an Hypersil ODS C18 HPLC column (250 mm × 4.6 mm, 5 μm particle size) by Thermo Fisher (Thermo Fisher Scientific, San Jose, CA, USA). The mobile phase consisted of eluent A, which was an aqueous solution of 30 mM ammonium acetate (adjusted at pH 5 with formic acid), and eluent B, MeOH. The flow rate was set at 0.2 mL/min. The elution gradient was as follows: started at 95% A and linearly decreased to 70% in 8 min and then to 50% in 5 min. This composition was held constant during 5 min. Then A was decreased to 20% in 15 min and to 0% in 14 min. This composition was held for 2 min and finally the eluent was returned to the initial conditions in 4.5 min (95% of A) and kept constant during 3.5 min. The column temperature was set at 25 °C, the injection volume was 10 μL, and the total running time was 58 min. The ESI parameters for the spectrometric detection were as follows: spray voltage, 4 kV; sheath gas (N₂, >95%), 35 (adimensional); auxiliary gas (N₂, >95%), 10 (adimensional); skimmer voltage, 18 V; capillary voltage, 35 V; tube lens voltage, 95 V; heater temperature, 305 °C; capillary temperature, 300 °C. The mass spectra were acquired employing four alternating acquisition functions: (1) full MS, ESI+, without fragmentation (the higher collisional dissociation (HCD) collision cell was switched off), mass resolving power 25000 full width at half-maximum (fwhm), scan time 0.25 s; (2) full MS, ESI- using the aforementioned settings, (3) all-ion fragmentation (AIF), ESI+, with fragmentation (HCD on, collision energy 30 eV), mass resolving power 10000 fwhm, scan time 0.10 s; (4) AIF, ESI- using the settings explained for (3). The mass range in the full scan experiments was set at m/z 50–1000.

The chromatograms and spectra were acquired using the external calibration mode, and they were processed using Xcalibur version 4.1, with Quanbrowser and Qualbrowser (Thermo Fisher Scientific, Les Ulis, France).

Preprocessing and Data Treatment. Raw data were imported into Compound Discoverer software (version 2.1, Thermo Scientific Scientific) for alignment and data gap filling. The resulting data set was imported as an XML file using Excel, version 2013 (Redmond, WA, USA), and then it was processed by SIMCA version 15.0 (Umetrics, Umeå, Sweden). The obtained variables, correlated to every detected compound, were named by the following unique concatenation for each one: [obtained high resolution mass–corresponding retention time].

Processing and Multivariate Data Analysis. Unsupervised analyses such as PCA and HCA were applied as exploratory data analysis for visualization of the analytical connections among samples. Outliers, potentially disturbing the proposed models, were detected using Hotelling's T^2 value; meanwhile, a DModX graphic allowed the consideration of additional possible moderate outliers.³¹ Detected outliers were excluded for further statistical analyses.

Supervised PLS-DA was used for the modeling of samples in relation to the settled classes Y . Further discriminant analyses based on selected variables derived from the VIP analyses were applied for classification/prediction models of rums.

The goodness of fit of the models was evaluated by the coefficient of determination (R^2X), the proportion of the variance of the response that is explained by the model (R^2Y), and the predictive ability parameter (Q^2Y). To estimate the ability of a model, and to avoid overfitting by estimating the number of significant components (NSCs) to be used, k -fold cross-validation (k -fold-CV) was applied.³²

For all supervised analyses, all rum samples were randomly divided into two sets: 4/5 of the samples were selected as the model set to build the statistical model, and the remaining 1/5 of the samples were employed as the external validation set. The validation of the established models was externally evaluated by calculating the percentage of samples correctly classified into their respective groups

(% CC). For determination of the accuracy of prediction, the root-mean-square error of cross-validation (RMSECV) was calculated.³³ The cross-validated ANOVA (CV-ANOVA) p value was calculated as a measure of significance for the observed group separation by the obtained variables.³⁴ Additionally, since some discriminant analysis methods have a high tendency to give overfitted models, DA-validations were carried out by a permutation test.³⁵

When a two-category model was applied, receiver operating characteristic (ROC) curve analysis was also employed, as it is recommended as a valuable tool for the validation of binary classifications.³⁶

Tentative Identification of Discriminant Compounds. A VIP variable selection method was applied to identify the most statistically significant variables. Putative identification was done on the basis of the exact mass composition, from the goodness of isotopic fit for the predicted molecular formula, MS/MS evaluation, and database matching (PubMed, Biocyc, ChEBI, HMDB, KEGG, Metabolights, and PubChem databases available from Compound Discoverer software). From the accurate mass, the molecular formula was derived using the built-in elemental composition generator provided by Xcalibur software. When several candidates were found for a given m/z value, metabolite identification was sorted by giving preference to those metabolites normally found in alcoholic beverages. If various potential candidates were obtained, Mass Frontier 7.0 software provided by Thermo Scientific was used to discard candidates on the basis of feasible explanation of the fragments/product ions obtained in MS and MS/MS spectra. The MS/MS spectra were also compared using free mass spectral databases such as Metlin³⁷ and MassBank.³⁸

Finally, when several potential candidates were found, the marker was simply annotated using its m/z –retention time label previously created for SIMCA processes.

RESULTS AND DISCUSSION

LC-HRMS Fingerprinting. Rum samples were analyzed using the LC-HRMS method described previously (LC-HRMS Analysis). The mass range in the full scan experiments was set at m/z 50–1000 in order to minimize the contribution of polymeric components (such as peptides or polysaccharides and lignin polymers),³⁹ so that they were also out of the range of the detectable compounds of the present study.

Chromatographic generic conditions are essential in metabolomic workflows in order to obtain representative results of each kind of matrix; therefore, specific chromatographic requirements were not optimized for the present study. Thus, chromatographic conditions previously applied in metabolomic studies carried out by our research group were applied.⁴⁰ Therefore, in this work, the chromatographic fingerprint was obtained by LC-HRMS using C₁₈ as the stationary phase (Hypersil ODS C₁₈ HPLC column) for the analysis of LC-amenable compounds in the nonvolatile fraction of rum. Both positive (ESI+) and negative (ESI-) ionization modes were applied to maximize the obtained information. A gradient elution of 30 mM ammonium acetate (pH 5 adjusted with formic acid) (solvent A) and methanol (solvent B) was used, covering a wide chromatographic separation and showing narrow and resolved peaks.

The sample analysis order was carried out according to a block design to guarantee their comparability and lack of potential analytical bias. After every brand analysis, a blank analysis was conducted. Sample analyses were carried out in triplicate.

Data Preprocessing. The LC-HRMS data obtained from the 30 different rums with both acquisition modes were processed by Compound Discoverer software. First, the following mass selector conditions for further processing were applied: the precursor selection uses the MS1 precursor

scan, the precursor mass is arranged between 50 and 1000 Da, and the S/N threshold is larger than 1.5.

The following processing procedure consisted of peak picking and integration, retention time alignment using an adaptive curve (mass tolerance of 5 ppm; maximum shift of 2 min), and unknown compound detection, where detected features with an intensity lower than 250000 and S/N higher than 3 in each set of data were extracted and merged into components according to ion adducts. The selection of these conditions allowed the efficient selection of data minimizing the loss of relevant analytical information. The software detected chromatographic peaks and the mass of corresponding compounds on the basis of a generated list of predicted compositions. Only the compounds detected in all three replicates of each sample (coefficient of variation of the signal intensity <25%) were used for further steps, avoiding the selection of false positives that could lead to errors in the classification process.

Finally, isotope and adduct peak grouping, unknown compound grouping (mass tolerance of 5 ppm; maximum shift of 2 min), and gap filling (mass tolerance of 5 ppm; S/N threshold of 1.5) were applied. Additionally, a blank subtraction (using blank samples prepared with mobile phase) was used for deleting interfering masses from the solvent or chromatographic system that could lead to multivariate analysis worsening. The processing settings are summarized in the workflow shown in Table S-1 in the Supporting Information.

A total of 90 observations (30 rums and their replicates) and 497 different peaks were processed and collected in the resulting XML file. The chosen categories for further rum classification were origin, barrel, age, and raw material, as previously described in Table 1.

Data Pretreatment. The obtained data set in XML format was imported into SIMCA software. Before the variables were subjected to PCA and PLS-DA, different mathematical pretreatments were applied.

Logarithmic transformation was used for all the detected compounds, giving values in a smaller range without masking the effect of small values within the data.²⁴ This conversion resulted in a significant decrease in root-mean-square error cross-validation (RMSECV) values in further discriminant analysis and a reduction in standard deviation (SD) on variable value from $SD = 588238.8$ to $SD = 0.5$, allowing an adequate variable range for the present metabolomic study.

Pareto scaling ($1/\sqrt{SD}$) was applied to PCA and PLS-DA in order to reduce the influence of intense peaks while emphasizing weaker peaks that may have more metabolic relevance (resulting mean variable value 4.35677).⁴¹

Multivariate Data Analysis. The resulting data set was submitted to multivariate analysis for rum classification by SIMCA software. PCA was performed using all of the rums to show a trend of intergroup separation on the scores plot of data obtained in both positive and negative modes. The PCA method was used to model, compress, and visualize data and to select outliers.¹³

First, Hotelling's T^2 method based on multivariate generalization of the 95% confidence interval allowed the exclusion from the model of three different potential outliers (Figure S-1A). In addition, two moderate outliers were excluded by the $DModX$ ($Dcrit = 0.05$) function, which provides measurements of the distance to the model (the relative standard deviation for each row) (Figure S-1B). Five outliers produced

by acquisition deviations, all from different brands, were detected. Those outliers were excluded for further statistical analyses, leaving 85 rum observations.

Then, the PCA model was automatically fitted using 85 rum observations (N) and 11 significant components (NSC). A score plot explaining 82% of the total variance was obtained (Figure S-2). For model fitting, the autofitting tool provided by SIMCA software was used while k -fold-CV ($k = 7$) was applied for determination of the optimal NSC number. This validation randomly divided the whole data set into 7 folds, and one of these was discarded at each time and used as a test set for evaluating the prediction ability of the model (Q^2).

Although no clear rum classification was observed for the selected categories in the PCA representation, a separation group is not always observable in PCA models because changes due to metabolomic differences are usually small in comparison to intra- and intersample variations.⁴² Nevertheless, good results, in terms of reproducibility of the replicate analyses, were obtained. Wide dispersion of the observations between different brands among the established and validated methods denoted an excellent metabolomics wealth for the present study.

Rum Classification Analysis. For further separation of the groups on the selected rum categories, different PLS-DA models were created for each of the classification groups: aging, raw material, and barrel. This kind of supervised analysis provides a deeper analysis of the main matrix characteristics of metabolic samples and differs from PCA by the addition of grouping variables that indicate the belonging category of samples. The rum aging parameter for PLS-DA was created by considering rums aged equal to or less than 7 years as young and rums aged for more than 7 years as old.

For all supervised analyses, all rum samples were randomly divided into two sets: 68 observations (4/5 of the samples) were selected as the model set to build the statistical model, and the remaining 17 observations (1/5 of the samples) were employed as the external validation set.

After fitting of the model and k -fold-CV, the validity of the discriminant analysis models was evaluated by studying the obtained R^2Y and Q^2Y . The difference between goodness of fit and goodness of prediction value should not exceed 0.3, and poor robustness of the model is usually suspected when that happens.⁸ Additionally, to avoid common overfitting of discriminant models, permutation tests were performed ($n = 200$). The permutation test rearranged the experiments randomly through changing the sort order of the classification variables, Y , and randomly assigning Q^2Y up to 200 times. Recommended values for good fitting of models have been described as R^2Y interception ~ 0.3 and Q^2Y interception < 0.05 .⁴³ Significant models were obtained for all of the finally selected categories, considering p values < 0.05 as suitable values for metabolomics studies. For each model, Fisher's probability was calculated for expressing the probability of the classification occurring by chance and was satisfied when $p < 0.05$ for 95% confidence. Evaluation values of the model validation are shown in Table S-2 in the Supporting Information,³⁴ showing suitable results.

First of all, a PLS-DA based on the classification of the rums by origin was carried out. As suspected, the model could not be correctly validated because the composition of rum is mostly dependent on the manufacturing process due to its complexity and variability, rather than the manufacturing country.⁸ It was corroborated by poor model validation features such as $R^2Y =$

0.774 and $Q^2Y = 0.600$ for an intentionally overfitted model (NSC = 10 and $N = 85$).

In addition, PLS-DA of the aging barrel used in the rum production processes showed clear group overlapping between French oak barrel class and mixed (French + American oak barrel) barrel. With the aim of improvement in the classification results, both groups were blended as "French-mix" (Figure 1).

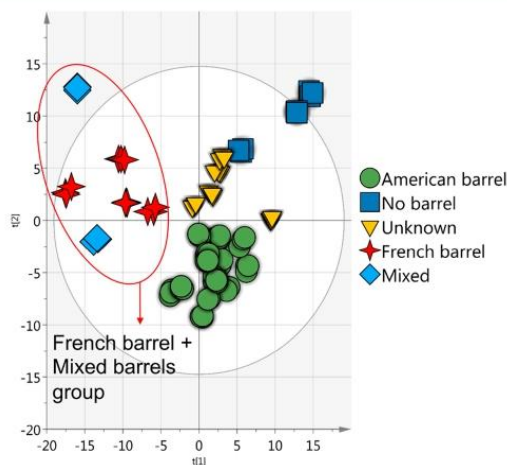


Figure 1. PLS-DA representation of 85 rums based on aging barrel and five classes (American barrel, no barrel, mixed barrel, French barrel, and unknown barrel).

The PLS-DA and its validation values obtained from rum classification on the basis of aging, raw material, and barrel are shown in Figure 2. Different categories were correctly separated after applying the PLS-DA, giving R^2Y values over 0.965 and Q^2Y values over 0.854. Those results revealed the stability of the mathematically designed model.

It can be observed in Table S-3 that, during model validation, samples were correctly classified using the proposed models except for one sample in the aging model, indicating the suitability of the selected experimental conditions to get valuable information for rum classification.

The results obtained by PLS-DA allowed a better understanding of those rums where no information was available, except for the origin, where no differences between countries were observed (Figure 2A), as has been indicated above. For raw material classification, clear differences between sugar cane and molasses were found ($R^2Y = 0.980$, $Q^2Y = 0.794$). Those rums for which the use of molasses or sugar cane is unknown are mainly classified within the molasses group, indicating the use of molasses during its fermentation. Three different rums, previously noted as white rums, were differentiated from the rest of the rums, suggesting the use of different or alternative raw materials for their preparation and showing good behavior of the proposed classification model (Figure 2B).

When the classification was performed according to barrel, for those unknown samples, clear graphical proximity indicated that American oak wood seems to be used when the manufacturers do not specify the type of barrel utilized (Figure 2C). On the other hand, for the French-mix group large

differences were found from the rest of the observations ($R^2Y = 0.961$, $Q^2Y = 0.908$).

For aging classification, clear differences were found among the three proposed classes: old rums (over 7 years old), young rums (less than or 7 years old), and not aged rums (Figure 2D). A linear trend depending on the rum aging was observed among the graphical representations ($R^2Y = 0.963$, $Q^2Y = 0.794$). The lower Q^2Y for the proposed classification model can be explained by the differences in legislation about age labeling. According to legislation from both the European Union and the United States, the age statement on the label needs to refer to the youngest rum in the bottle. However, in other countries, it can refer to the oldest rum.⁸

Model Validation-Prediction Ability. External validation was carried out by evaluating the prediction capability and performance of the designed models, calculating the percentage of samples correctly classified into their respective groups (% CC). Results of the classification of the different PLS-DA models according to raw material, barrel, and aging are shown in Table 2. With regard to the model prediction, suitable global classification, rating from 94 to 100%, was achieved for the validation sample sets of rums.

Additionally, ROC analysis of aging based on young and old groupings showed excellent results on the basis of the area under the curve values (AUC) equal to $AUC = 0.99$ as shown in Figure S-3, showing an excellent behavior for this rum binary classification.

According to the obtained results, differentiation of rums on the basis of the selected categories (age, raw material, and barrel) considered in the present research was demonstrated.

Elucidation of Markers and Biological Meaning. After model design and validation, a VIP variable selection method was applied for the determination and identification of the most discriminant and characteristic compounds in the selected rum categories. The dimensional reduction of the total detected variables allowed a better and more comprehensive elucidation of the most important compounds of the studied rum samples. It also contributed to improving the understanding of the chemical meaning of this kind of spirit beverage.

For the identification of the most discriminant compounds obtained from VIP analysis, Compound Discoverer 2.1 software (Thermo Scientific) was applied. From the initial 497 detected compounds, 10 different compounds were selected, showing a high VIP value ($VIP > 1.5$), whereas other 6 expected compounds in alcoholic beverages were identified due to the high concentration at which they were present, demonstrating the elucidation ability of the proposed methodology.

For a compound formula generation, the maximum numbers of possible C, H, Cl, N, O, P, and S atoms allowed for the built-in elemental composition generator were 90, 190, 2, 10, 15, 2, and 2, respectively. These values were selected in order to ensure the elucidation of expected compounds in rums such as sugars and furfural derivatives. Mass tolerance for MS identification of the exact mass of potential candidates was 5 ppm. Only for one compound, L-proline, it was necessary to increase the mass tolerance in order to obtain potential candidates. For the spectral data selection, the preferred ions were selected on the basis of the mobile phase used in the present study and results were ranked by the spectral similarity score between theoretical and measured isotope patterns

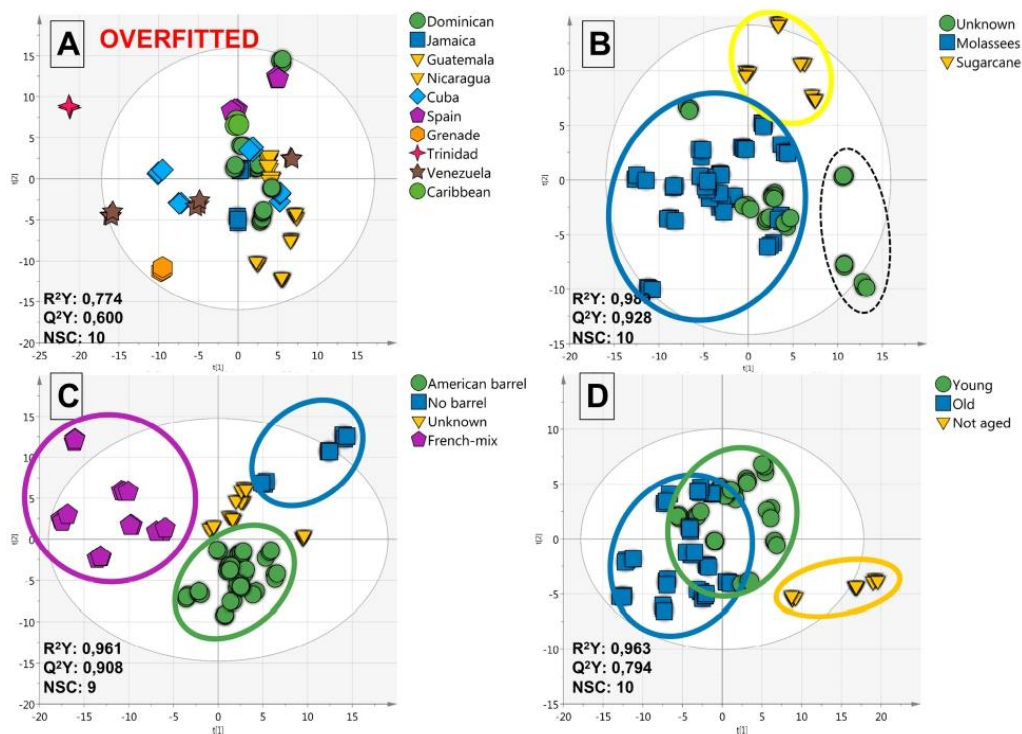


Figure 2. PLS-DA representation showing the classification of 85 rums on the basis of (A) origin, (B) raw material, (C) aging barrel, and (D) aging period.

Table 2. Classification Results (CC%) for Rum Validation Set in the Proposed Models

model	total of sampling set	number of categories	total of correct classified	correctly classified (CC %)
raw material	17	3	16	94
barrel	17	4	17	100
aging	17	3	17	100

(pattern cov, %). Details of the Compound Discoverer steps and parameters are summarized in Table S-1.

After exact mass composition determination, a library searching was carried out by comparing both the exact formula and predicted formula obtained with seven different available databases: PubMed, Biocyc, ChEBI, HMDB, KEGG, Metabolics, and PubChem. The database selection was based on their suitable applications for metabolite identification from mass spectrometry based studies.²⁹

The generated list of potential metabolites for each PLS-DA model is shown in Table 3, including some characteristic parameters. In addition, matched covalence indicates the matched intensity percentage of the theoretical pattern for the proposed compounds. Mass Frontier 7.0 software was used to select candidates on the basis of the presence/absence of theoretical fragments in the experimental MS/MS spectra (theoretical MS/MS fragments). For 1-deoxy-1-nitro-D-mannitol, 4-guanidinobutanoic acid, and L-(+)-tartaric acid diethyl ester, no confirmation by fragments was obtained because

these compounds were found at low intensities, and adequate MS/MS information for the present structural elucidation could not be attained.

All of the selected compounds showed how the previous understanding of the rum elaboration process is necessary during the elucidation in order to be focused on the most plausible candidates. Thus, acid-catalyzed decomposition of those carbohydrates (cellulose and hemicellulose) from wood was demonstrated to be clearly correlated with the classification of rums on the basis of aging and barrel categories (Figure 3),⁴⁴ leading to the generation of furfurals, such as 5-(hydroxymethyl)furfural (Table 3).

Among the found compounds, different sugar derivatives were identified as the most discriminant compounds for the selected categories for rum classification with high VIP values. In both molasses and sugar cane juice raw materials used for rum elaboration, the main components are sugars such as sucrose, glucose, and fructose and non-sugar compounds including citric acid and oxalic acid.⁴⁵ Fructosylglycine, which has the greatest discriminant power of the proposed models after VIP analysis (VIP score 2.04), is a monosaccharide derivative, and MS/MS elucidation leads to the identification of its two different main fragments, as is shown in Figure 4. Other sugar derivatives as 1-deoxy-1-nitro-D-mannitol (VIP score 1.71), methyl 4,6-dideoxy-4-(3-deoxy-L-glycero-tetramido)-2-O-methyl- α -D-mannopyranoside (VIP score 1.58) and methyl 6-deoxy- α -D-glucopyranoside (VIP score 1.51) were found to be compounds with high discriminant ability for the classification of rums. Although no high discrimination ability

Table 3. Selected Components Identified as the Most Discriminant Compounds in Rums by LC-Orbitrap-MS

retention time (min)	metabolite ID	metabolite formula	assigned adduct	theor mol wt	measd mol wt	$\Delta(\text{mass})$ (ppm)	pattern cov (%)	theoretical ms/ms fragment	VIP value (VIP class)
9.199	fructosylglycine	C ₈ H ₁₃ NO ₇	[M + H] ⁺	237.08485	237.08483	0.09	98.43	133.04954 115.02639	2.04 (barrel)
9.001	1-deoxy-1-nitro-D-mannitol	C ₈ H ₁₃ NO ₇	[M + H] ⁺	211.06920	211.06938	0.84	98.27		1.71 (age) 1.66 (raw material)
10.726	L-proline	C ₅ H ₉ NO ₂	[M + H] ⁺ [M + Na] ⁺	115.06333	115.06411	6.76	100.00	70.06513	1.83 (raw material) 1.63 (age)
11.797	4-guanidinobutanoic acid	C ₅ H ₁₁ N ₃ O ₂	[M + H] ⁺	145.08513	145.08532	1.31	99.61		1.98 (raw material)
25.311	1 <i>H</i> -cyclopenta[<i>c</i>]furan-1,3(4 <i>H</i>)-dione-5,6-dihydro	C ₇ H ₈ O ₄	[M + H] ⁺	156.04225	156.04237	0.73	99.23	115.03897 129.01824	1.52 (raw material) 1.50 (barrel)
27.520	L-(+)-tartaric acid diethyl ester	C ₈ H ₁₄ O ₆	[M + Na] ⁺ [M + H] ⁺	206.07904	206.07906	0.09	99.90		1.91 (raw material)
35.911	methyl 4,6-dideoxy-4-(3-deoxy-L-glycero-tetramido)-2-O-methyl- α -D-mannopyranoside	C ₁₂ H ₂₃ NO ₇	[M + H] ⁺	293.14746	293.14752	0.21	98.04	133.04954 115.03897	1.58 (raw material)
9.200	(-)-threo-isodihomocitric acid	C ₈ H ₁₂ O ₇	[M + Na] ⁺ [M + H] ⁺ [M + NH ₄] ⁺	220.05830	220.05817	0.61	100.00	157.04954	1.57 (barrel)
16.473	methyl 6-deoxy- α -D-glucopyranoside	C ₇ H ₁₄ O ₅	[M + NH ₄] ⁺	178.08412	178.08435	1.30	98.66	73.02841	1.51 (barrel)
23.077	hydroxymethylfurfural	C ₆ H ₆ O ₃	[M + Na] ⁺ [M + H] ⁺ [M + NH ₄] ⁺	126.03169	126.03201	2.51	100.00	81.03349 109.02841	1.98 (age) 1.79 (barrel)
9.718	hexose	C ₆ H ₁₂ O ₆	[M + NH ₄] ⁺	180.06338	180.06277	3.43	100.00	85.02933 145.04954	<1.5
9.770	galactosamine	C ₆ H ₁₃ NO ₃	[M + H] ⁺	179.07938	179.07947	0.52	100.00	127.03897 145.04954	<1.5
9.846	D-gluconic acid	C ₆ H ₁₂ O ₇	[M - H] ⁻	196.05830	196.05719	4.68	100.00	89.02442	<1.5
30.065	(+)-menthol acetate	C ₁₂ H ₂₂ O ₂	[M - H] ⁻	198.16199	198.16125	3.70	99.20	197.15361	<1.5
30.066	homovanillyl alcohol	C ₉ H ₁₂ O ₃	[M + H] ⁺	168.07864	168.07875	0.65	99.39	109.06479 154.06245	<1.5
30.197	5-[(4-methoxyphenoxy)methyl]-2-furoic acid	C ₁₃ H ₁₂ O ₅	[M + H] ⁺	248.06847	248.06868	0.86	98.44	81.03349 109.02841	<1.5

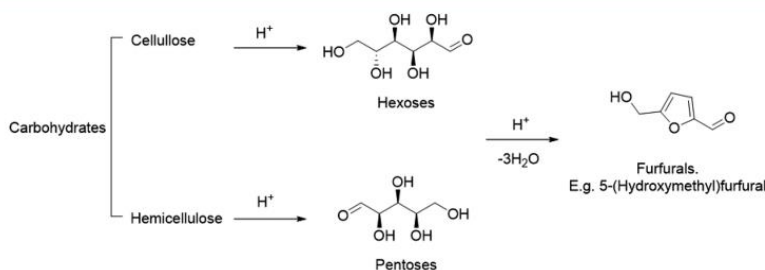


Figure 3. Acid-catalyzed decomposition of carbohydrates via dehydroxilation. Adapted from ref 44.

was found for all of the expected compounds, other saccharide compounds such as hexose and galactosamine were found at high concentrations in all of the analyzed rums. Additionally, (-)-threo-isodihomocitric acid (VIP score 1.57) was also identified as a derivate of citric acid (Table 3).

Furans such as 5-hydroxymethylfurfural (HMF), 5,6-dihydro-1*H*-cyclopenta[*c*]furan-1,3(4*H*)-dione, and 5-[(4-methoxyphenoxy)methyl]-2-furoic acid, produced through acid-catalyzed dehydration and hydrolysis of hexose sugars, were also detected at high concentrations.

Although MS/MS confirmation of 4-guanidinobutanoic acid was not possible due to its low concentration, guacyl-type

compounds were expected due to barrel fabrication processes resulting in lignin thermodegradation involving the formation of aromatic hydrocarbons, phenolics, hydroxyphenolics, and guaiacyl-type compounds.³⁹ 4-Guanidinobutanoic acid has also been related to metabolic reactions in yeast using *Saccharomyces cerevisiae* enzyme.⁴⁶

Different characteristic compounds of rum's flavor have been found and identified by the proposed methodology. L-(+)-Tartaric acid diethyl ester, also known as diethyl tartrate, is a common flavoring ingredient typically present in alcoholic beverages.⁴⁷ Vanillin and menthol derivate compounds were

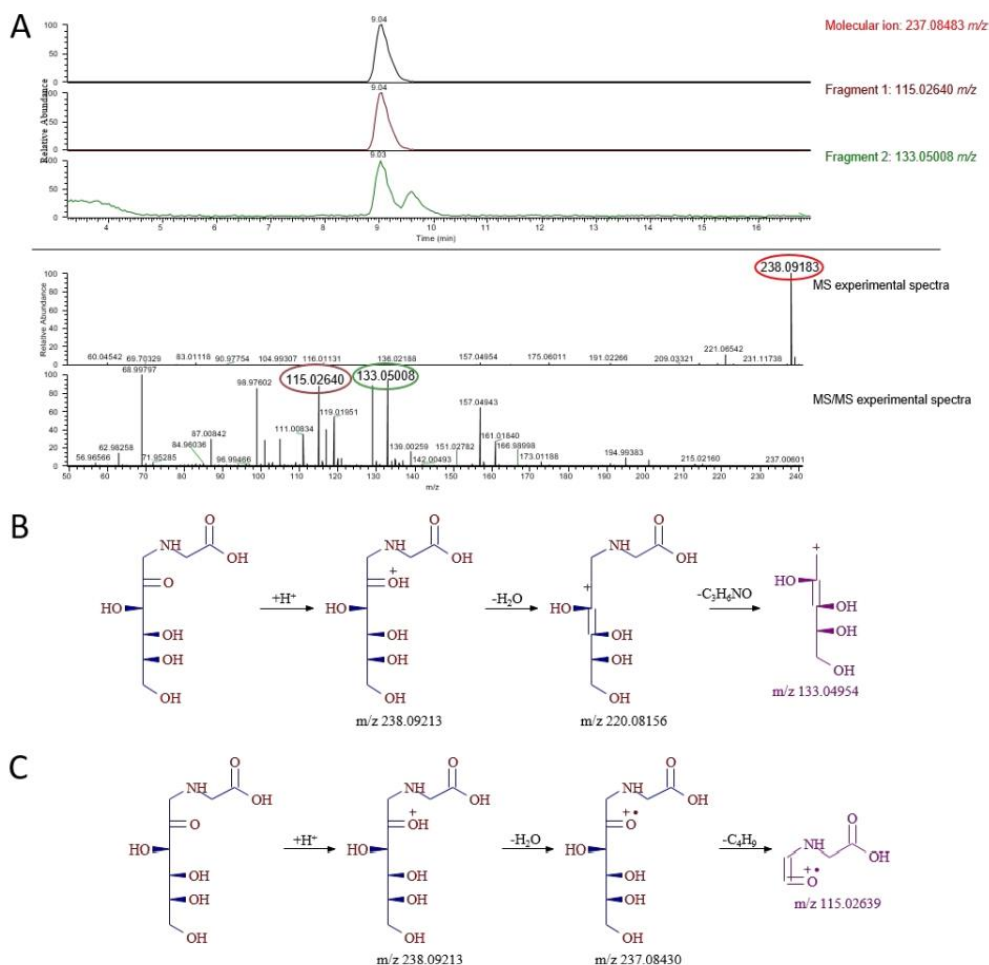


Figure 4. (A) Fructosylglycine fragment identification based on its chromatographic profile and MS and MS/MS spectra. (B, C) Structural elucidation of 115.03740 m/z and 133.05012 m/z fragments obtained for fructosylglycine.

detected and identified as homovanillyl alcohol and (+)-menthol acetate, respectively.⁴⁸

Another interesting studied compound is L-proline, which is an osmoprotectant and a sweet amino acid that could protect yeast cells from damage by freezing, desiccation, or oxidative stress. In addition, *S. cerevisiae* synthesizes L-proline from L-glutamate via glutamyl type enzymatic pathways in the fermentation processes.⁴⁹

In conclusion, a nontargeted metabolomic approach of golden rums performed by LC-Orbitrap-MS was successfully developed for the discrimination of the samples on the basis of their production characteristics. Fermentation barrel, raw material, and aging of rums were selected as classification categories showing excellent fitting (R^2Y from 0.963 to 0.980) and predictive abilities (Q^2Y from 0.794 to 0.908) performed by PLS-DA. For those samples where their labeling did not provide all the information about their method of preparation, the results obtained have sorted out this issue. Most of the rums with insufficient labeling have been developed by the use of molasses in the fermentation process followed by American

oak barrels for the maturation stage. The proposed models have been evaluated to be applied as rum classification tools by analyzing a validation set, demonstrating excellent results of approximately 100% success.

The results confirmed a strong effect of different sugar derivatives such as fructosylglycine (VIP_{barrel} score 2.04) and 1-deoxy-1-nitro-D-mannitol (VIP_{raw material} score 1.66) and furfural derivatives (e.g., 5-hydroxymethylfurfural, VIP_{Age} score 1.79) in the selected categories. In addition, the presence of flavouring compounds such as L-(+)-tartaric acid diethyl ester (VIP score 1.91) can be decisive when it comes to distinguishing rums.

Therefore, promising data have been obtained for further studies focused on the markers selected in this paper (validation of a target quantitative method for suitable classification methods on target acquisition mode), the application of the proposed metabolomics approach in different scenarios (such as golden/white rum comparisons) and data fusion for multiplatform classification of rums. Thus, the metabolomic approach is a powerful tool for rum

classification, providing suitable information regarding the chemical markers related to aging, barrel, or raw material, providing a deeper understanding of the metabolomic chemistry of golden rums, as well as to avoid potential fraud of this “quality premium” product.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jafc.8b05622.

Data analysis workflow in Compound Discoverer 2.1 software, validation data obtained for each classification data, misclassification table, Hotelling's T^2 representation, PCA representation of the 85 rums and 82% of variance, and ROC representation for the evaluation of binary classification (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*A.G.F.: tel, +0034950015985; fax, +003495005008. e-mail, agarrido@ual.es.

ORCID

Antonia Garrido French: 0000-0002-7904-7842

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS USED

AUC, area under the curve; CC, correctly classified; DA, discriminant analysis; GC-MS, gas chromatography–mass spectrometry; HCA, hierarchical cluster analysis; LC-HRMS, liquid chromatography coupled with high-resolution mass spectrometry; NSC, number of significant components; OPLS-DA, orthogonal partial least-squares discriminant analysis; PCA, principal component analysis; PLS-DA, partial least-squares discriminant analysis; RMSECV, root-mean-square error of cross-validation; ROC, receiver operating characteristic; VIP, variable importance in projection.

■ REFERENCES

- (1) Golden Rum: The Whisky Exchange; <https://www.thewhiskyexchange.com/c/518/golden-rum> (accessed Apr 17, 2018).
- (2) Reche, R. V.; Neto, A. F. L.; Da Silva, A. A.; Galinaro, C. A.; De Osti, R. Z.; Franco, D. W. Influence of Type of Distillation Apparatus on Chemical Profiles of Brazilian Cachacas. *J. Agric. Food Chem.* **2007**, *55* (16), 6603–6608.
- (3) Franitza, L.; Granvogl, M.; Schieberle, P. Influence of the Production Process on the Key Aroma Compounds of Rum: From Molasses to the Spirit. *J. Agric. Food Chem.* **2016**, *64* (47), 9041–9053.
- (4) European Commission. EC Regulation 110/2008 of the European Parliament and of the Council on the Definition, Description, Presentation, Labelling and the Protection of Geographical Indications of Spirit Drinks. *Off. J. Eur. Union* **2008**, *L39*, 16–54.
- (5) Lea, A. G. H.; Piggott, J. *Fermented Beverage Production*; Springer Science & Business Media: 2012.
- (6) Navarro, M.; Kontoudakis, N.; Gómez-Alonso, S.; García-Romero, E.; Canals, J. M.; Hermosín-Gutiérrez, I.; Zamora, F. Influence of the Volatile Substances Released by Oak Barrels into a Cabernet Sauvignon Red Wine and a Discolored Macabeo White Wine on Sensory Appreciation by a Trained Panel. *Eur. Food Res. Technol.* **2018**, *244* (2), 245–258.
- (7) Wiśniewska, P.; Sliwińska, M.; Dymerski, T.; Wardencki, W.; Namieśnik, J. Application of Gas Chromatography to Analysis of Spirit-Based Alcoholic Beverages. *Crit. Rev. Anal. Chem.* **2015**, *45* (3), 201–225.
- (8) Belmonte-Sánchez, J. R.; Gherghel, S.; Arrebola-Liébanas, J.; González, R. R.; Vidal, J. L. M.; Parkin, I.; French, A. G. Rum Classification Using Fingerprinting Analysis of Volatile Fraction by Headspace Solid Phase Microextraction Coupled to Gas Chromatography–Mass Spectrometry. *Talanta* **2018**, *187* (May), 348–356.
- (9) Franitza, L.; Nicolotti, L.; Granvogl, M.; Schieberle, P. Differentiation of Rums Produced from Sugar Cane Juice (Rhum Agricole) from Rums Manufactured from Sugar Cane Molasses by a Metabolomics Approach. *J. Agric. Food Chem.* **2018**, *66* (11), 3038–3045.
- (10) Oller-Ruiz, A.; Viñas, P.; Campillo, N.; Fenoll, J.; Hernández-Córdoba, M. Triple Quadrupole Mass Spectrometry with Liquid Chromatography and Dispersive Liquid-Liquid Microextraction for the Determination of Monoterpenes in Alcoholic Drinks. *Food Anal. Methods* **2017**, *10* (11), 3615–3622.
- (11) Barnaba, C.; Dellacassa, E.; Nicolini, G.; Nardin, T.; Malacarne, M.; Larcher, R. Identification and Quantification of 56 Targeted Phenols in Wines, Spirits, and Vinegars by Online Solid-Phase Extraction - Ultrahigh-Performance Liquid Chromatography - Quadrupole-Orbitrap Mass Spectrometry. *J. Chromatogr. A* **2015**, *1423*, 124–135.
- (12) Regalado, E. L.; Tolle, S.; Pino, J. A.; Winterhalter, P.; Menendez, R.; Morales, A. R.; Rodríguez, J. L. Isolation and Identification of Phenolic Compounds from Rum Aged in Oak Barrels by High-Speed Countercurrent Chromatography/High-Performance Liquid Chromatography-Diode Array Detection-Electrospray Ionization Mass Spectrometry and Screening for Antioxidant. *J. Chromatogr. A* **2011**, *1218* (41), 7358–7364.
- (13) Wiśniewska, P.; Dymerski, T.; Wardencki, W.; Namieśnik, J. Chemical Composition Analysis and Authentication of Whisky. *J. Sci. Food Agric.* **2015**, *95* (11), 2159–2166.
- (14) Kew, W.; Goodall, I.; Clarke, D.; Uhrin, D. Chemical Diversity and Complexity of Scotch Whisky as Revealed by High-Resolution Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 200–213.
- (15) Roullier-Gall, C.; Signoret, J.; Hemmler, D.; Witting, M. A.; Kanawati, B.; Schäfer, B.; Gougeon, R. D.; Schmitt-Kopplin, P. Usage of FT-ICR-MS Metabolomics for Characterizing the Chemical Signatures of Barrel-Aged Whisky. *Front. Chem.* **2018**, *6* (2), 1–11.
- (16) Slaghenauf, D.; Franc, C.; Mora, N.; Marchand, S.; Perello, M. C.; Revel, G. de. Quantification of Three Galloylglucoside Flavour Precursors by Liquid Chromatography Tandem Mass Spectrometry in Brandies Aged in Oak Wood Barrels. *J. Chromatogr. A* **2016**, *1442*, 26–32.
- (17) Ellis, D. I.; Muhamadali, H.; Allen, D. P.; Elliott, C. T.; Goodacre, R. A Flavour of Omics Approaches for the Detection of Food Fraud. *Curr. Opin. Food Sci.* **2016**, *10*, 7–15.
- (18) Luginbühl, M.; Schröck, A.; König, S.; Schürch, S.; Weinmann, W. Determination of Fatty Acid Ethyl Esters in Dried Blood Spots by LC–MS/MS as Markers for Ethanol Intake: Application in a Drinking Study. *Anal. Bioanal. Chem.* **2016**, *408* (13), 3503–3509.
- (19) Gallart-Ayala, H.; Kamleh, M. A.; Hernández-Cassou, S.; Saurina, J.; Checa, A. Ultra-High-Performance Liquid Chromatography–high-Resolution Mass Spectrometry Based Metabolomics as a Strategy for Beer Characterization. *J. Inst. Brew.* **2016**, *122* (3), 430–436.
- (20) Teodoro, J. A. R.; Pereira, H. V.; Sena, M. M.; Piccin, E.; Zacca, J. J.; Augusti, R. Paper Spray Mass Spectrometry and Chemometric Tools for a Fast and Reliable Identification of Counterfeit Blended Scottish Whiskies. *Food Chem.* **2017**, *237*, 1058–1064.
- (21) Li, J.; Jiang, K.; Wang, L.-J.; Yin, G.; Wang, J.; Wang, Y.; Jin, Y.-B.; Li, Q.; Wang, T.-J. HPLC-MS/MS Determination of Flavonoids in *Gleditsia Spina* for Its Quality Assessment. *J. Sep. Sci.* **2018**, *41* (8), 1752–1763.
- (22) Canas, S. Phenolic Composition and Related Properties of Aged Wine Spirits: Influence of Barrel Characteristics. A Review. *Beverages* **2017**, *3* (4), 55.

- (23) Efenberger-Szmechtyk, M.; Nowak, A.; Kregiel, D. Implementation of Chemometrics in Quality Evaluation of Food and Beverages. *Crit. Rev. Food Sci. Nutr.* **2018**, *58* (10), 1747–1766.
- (24) Lubes, G.; Goodarzi, M. Analysis of Volatile Compounds by Advanced Analytical Techniques and Multivariate Chemometrics. *Chem. Rev.* **2017**, *117* (9), 6399–6422.
- (25) Callao, M. P.; Ruisánchez, I. An Overview of Multivariate Qualitative Methods for Food Fraud Detection. *Food Control* **2018**, *86*, 283–293.
- (26) Mung, D.; Li, L. Applying Quantitative Metabolomics Based on Chemical Isotope Labeling LC-MS for Detecting Potential Milk Adulterant in Human Milk. *Anal. Chim. Acta* **2018**, *1001*, 78–85.
- (27) Liu, Y. J.; Tran, T.; Postma, G.; Buydens, L. M. C.; Jansen, J. Estimating the Number of Components and Detecting Outliers Using Angle Distribution of Loading Subspaces (ADLS) in PCA Analysis. *Anal. Chim. Acta* **2018**, *1020*, 17–29.
- (28) Farag, M. A.; Khattab, A. R.; Ehrlich, A.; Kropf, M.; Heiss, A. G.; Wessjohann, L. A. Gas Chromatography/Mass Spectrometry-Based Metabolite Profiling of Nutrients and Antinutrients in Eight Lens and Lupinus Seeds (Fabaceae). *J. Agric. Food Chem.* **2018**, *66* (16), 4267–4280.
- (29) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal. Chim. Acta* **2016**, *914*, 17–34.
- (30) Ji, H. G.; Lee, Y. R.; Lee, M. S.; Hwang, K. H.; Park, C. Y.; Kim, E. H.; Park, J. S.; Hong, Y. S. Diverse Metabolite Variations in Tea (*Camellia Sinensis* L.) Leaves Grown under Various Shade Conditions Revisited: A Metabolomics Study. *J. Agric. Food Chem.* **2018**, *66* (8), 1889–1897.
- (31) Marini, F. Classification Methods in Chemometrics. *Curr. Anal. Chem.* **2010**, *6* (1), 72–79.
- (32) Triba, M. N.; Le Moyec, L.; Amathieu, R.; Goossens, C.; Bouchemal, N.; Nahon, P.; Rutledge, D. N.; Savarin, P. PLS/OPLS Models in Metabolomics: The Impact of Permutation of Dataset Rows on the K-Fold Cross-Validation Quality Parameters. *Mol. BioSyst.* **2015**, *11* (1), 13–19.
- (33) Xi, B.; Gu, H.; Baniasadi, H.; Raftery, D. Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data. In *Molecular Analysis and Genome Discovery*; Wiley: Chichester, U.K., 2014; pp 333–353.
- (34) Eriksson, L.; Trygg, J.; Wold, S. CV-ANOVA for Significance Testing of PLS and OPLS® Models. *J. Chemom.* **2008**, *22* (11–12), 594–600.
- (35) Gonzalez-Riano, C.; Sanz-Rodríguez, M.; Escudero-Ramirez, J.; Lorenzo, M. P.; Barbas, C.; Cubelos, B.; García, A. Target and Untargeted GC-MS Based Metabolomic Study of Mouse Optic Nerve and Its Potential in the Study of Neurological Visual Diseases. *J. Pharm. Biomed. Anal.* **2018**, *153*, 44–56.
- (36) Broadhurst, D. I.; Kell, D. B. Statistical Strategies for Avoiding False Discoveries in Metabolomics and Related Experiments. *Metabolomics* **2007**, *2* (4), 171–196.
- (37) Metlin Scripps; https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage (accessed Sep 2, 2018).
- (38) MassBank. High Quality Mass; DOI: 10.1002/jms.1777 (accessed Sep 2, 2018).
- (39) Le Floch, A.; Jourdes, M.; Teissedre, P. L. Polysaccharides and Lignin from Oak Wood Used in Cooperage: Composition, Interest, Assays: A Review. *Carbohydr. Res.* **2015**, *417*, 94–102.
- (40) López-Gutiérrez, N.; Romero-González, R.; Vidal, J. L. M.; Frenich, A. G. Quality Control Evaluation of Nutraceutical Products from Ginkgo Biloba Using Liquid Chromatography Coupled to High Resolution Mass Spectrometry. *J. Pharm. Biomed. Anal.* **2016**, *121*, 151–160.
- (41) Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2012**, *1* (1), 92–107.
- (42) Van, Q. N. *Current NMR Strategies for Biomarker Discovery*; Elsevier: 2013.
- (43) Esteban-Fernandez, A.; Ibañez, C.; Simó, C.; Bartolome, B.; Moreno-Arribas, M. V. An UHPLC-TOF MS Metabolomic Approach to Study the Impact of Moderate Red Wine Consumption on Urinary Metabolome. *J. Proteome Res.* **2018**, *17*, 1624.
- (44) O'Hara, I. M.; Mundree, S. G. *Sugarcane-Based Biofuels and Bioproducts*; **2016**, 1.
- (45) El-Gendy, N. S.; Madian, H. R.; Amr, S. S. A. Design and Optimization of a Process for Sugarcane Molasses Fermentation by *Saccharomyces Cerevisiae* Using Response Surface Methodology. *Int. J. Microbiol.* **2013**, *2013*, 1.
- (46) 4-Guanidinobutanoic acid (CHEBI: 15728); <http://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:15728> (accessed Jul 26, 2018).
- (47) Rayne, S.; Forest, K. Carboxylic Acid Ester Hydrolysis Rate Constants for Food and Beverage Aroma Compounds. *Flavour Fragrance J.* **2016**, *31* (5), 385–394.
- (48) Johnson, A. J.; Heymann, H.; Ebeler, S. E. Volatile and Sensory Profiling of Cocktail Bitters. *Food Chem.* **2015**, *179*, 343–354.
- (49) Takagi, H.; Takaoka, M.; Kawaguchi, A.; Kubo, Y. Effect of L-Proline on Sake Brewing and Ethanol Stress in *Saccharomyces Cerevisiae*. *Appl. Environ. Microbiol.* **2005**, *71* (12), 8656–8662.

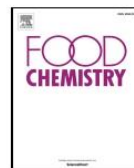


2.5 Aportación científica 3. $^1\text{H-NMR}$ and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis

J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido Frenich. Food Chemistry. 317 (2020) 126363.

DOI: 10.1016/j.foodchem.2020.126363

CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA AL ESTUDIO DE RON



¹H NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis

José Raúl Belmonte-Sánchez¹, Roberto Romero-González², José Luis Martínez Vidal³, Francisco Javier Arrebola⁴, Antonia Garrido Frenich^{5,*}

Department of Chemistry and Physics, Research Centre CIAIMBITAL, Agrifood Campus of International Excellence, ceiA3, University of Almería, Ctra. Sacramento, s/n, 04120, Almería, Spain



ARTICLE INFO

Keywords:

Rum
NMR spectroscopy
Metabolomics
Multivariate analysis
Data-fusion

ABSTRACT

¹H NMR spectroscopy combined with chemometrics was applied for the first time for golden rum classification based on several factors as fermentation barrel, raw material, distillation method and aging. Principal component analysis (PCA) was used to assess the overall structure, and partial least square discriminant analysis (PLS-DA) was carried out for the analytical discrimination of rums. Additionally, data-fusion of ¹H NMR and chromatographic techniques (gas and liquid chromatography) coupled to mass spectrometry was applied to provide more accurate knowledge about rums. This approach provided a classification of samples with lower error rate than the one obtained by the use of a single technique (spectroscopic or chromatographic). The results showed that ¹H NMR spectroscopy is an appropriate technique for the suitable classification of > 95.5% of the samples. When data fusion methodology of spectroscopic and spectrometric data was performed, the prediction efficiency can reach 100% of the samples.

1. Introduction

Nuclear magnetic resonance (NMR) is a suitable tool in food control, considering that it is a robust, quick, reproducible, non-destructive and relatively easy to use analytical platform that does not require laborious sample preparation (Ralli et al., 2018). In addition, NMR technique allows for monitoring different classes of chemical compounds, within the same experiment, providing a general overview of the composition of the studied matrix (Marini, 2010).

During the last few years, NMR-based metabolomics, especially ¹H NMR acquisition, has played a principal role in the characterization of complex food matrices, and it has been largely used in metabolomic studies of foodstuffs, allowing geographical discrimination as well as its quality and authentication assessment (Fortunato, Rocha, & Sheen, 2018). NMR-based metabolomics has been used to classify different types of matrices, such as beer (da Silva, Flumignan, Tininis, Pezza, & Pezza, 2019), whiskey (Monakhova et al., 2011), grape marc spirits (Fotakis & Zervou, 2016) and wine (Gougeon et al., 2018). Although

previous metabolomics studies for golden rums have been reported by LC coupled to mass spectrometry (LC-MS) (Belmonte-Sánchez, Romero-González, Arrebola, Vidal, & Garrido Frenich, 2019) and gas chromatography coupled to MS (GC-MS) (Belmonte-Sánchez et al., 2018), up to our knowledge, NMR-based untargeted metabolomics has never been applied for golden rums comparison.

Rum preparation is based on different complex stages. First, the raw material used for fermentation by the action of yeast enzymes, is selected between sugarcane or molasses, and ethanol is produced by alcoholic fermentation in large vats. The obtained mash is distilled using heat to obtain a liquor with a high content of ethanol. Both copper pot stills and stainless-steel columns can be used. To get the desired alcohol volume of 37.5–40.0%, dilution of the obtained distillate can be done, before or after the aging step, using pure demineralized water. A minimum of 37.5% v/v alcohol is required for any commercial rum according to European regulation (European Commission, 2008). Finally, the obtained solution is aged in oak barrels previously used for other alcoholic beverages (Lea & Piggott, 2012). American white oak

* Corresponding author.

E-mail address: agarrido@ual.es (A. Garrido Frenich).

¹ José Raúl Belmonte-Sánchez: 0000-0003-4683-5903.

² Roberto Romero-González: 0000-0002-2505-2056.

³ José Luis Martínez Vidal: 0000-0003-0655-2597.

⁴ Francisco Javier Arrebola: 0000-0003-0760-6008.

⁵ Antonia Garrido Frenich: 0000-0002-7904-7842.

(*Q. alba*) and French red oak (*Q. robur* and *Q. petraea*) are the most commonly oak woods used (Navarro et al., 2018). To achieve particular sensory characteristics, blending can be an optional step. Specific organoleptic properties can be obtained by mixing rums of different ages, and the addition of flavourings, sugars or other sweetening products is forbidden to be considered a spirit drink without extra addition (Coldea, Mudura, & Socaciu, 2017). As it can be seen, rum classification represents an analytical challenge due to the complexity and variability of their preparation processes. Additionally, manufacturers did not always provide information related to different rum's properties as aging, raw material or barrel, most of the cases attributed to those brands of lower price. This lack of information hampers the establishment of appropriate quality models of rums, in order to improve their knowledge in terms of composition and properties.

The implementation of chemometric methods combined with the potentiality of NMR spectroscopy allows for obtaining comprehensive fingerprints and therefore, it can be used in metabolomics researches. ¹H NMR experiment (1 dimensional nuclear Overhauser effect spectroscopy (NOESY)) with water and ethanol suppression followed by multivariate analysis is a powerful tool for the discrimination of alcoholic beverages and for their quality control (Longobardi et al., 2017; Monakhova et al., 2011). In matrices with high content of major components (as ethanol in rums), the ¹³C satellite signals suppression is also important to obtain an adequate dynamic range, which allows the correct acquisition of low-concentration metabolites (Bharti & Roy, 2012).

In the last years, diverse supervised and unsupervised methods, such as principal component analysis (PCA), hierarchical clustering analysis (HCA), partial least square discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA) have already been used applying different analytical techniques (NMR, HPLC-MS, GC-MS, Raman spectroscopy, etc) for metabolomic analysis of food and beverages quality (Callao & Ruisánchez, 2018; Fortunato et al., 2018; Mung & Li, 2018). The unsupervised PCA method is the most widely used data analysis tool for NMR based metabolomics to obtain a preliminary evaluation of the data, finding trends and grouping within the dataset (Farag et al., 2018; Razali et al., 2018). Supervised techniques can be applied for a better understanding of known data structures, looking for patterns of the analysed rums. Prediction of new data using supervised methods (e.g. PLS-DA, LDA, OPLS-DA) can be applied for targeted rum classification (Hoyos Ossa, Gil-Solsona, Peñuela, Sancho, & Hernández, 2018; Longobardi et al., 2017).

In order to improve the obtained metabolomics results, a further step can be done by combining the outputs of multiple instrumental sources. Low-level fusion classification can be obtained unifying all the data into a unique matrix containing all the information from the different instruments before building the model. Feature level fusion (also called mid-level fusion) can be applied by concatenating the obtained principal components of the individually built models into a single array that is applied for multivariate classification (Biancolillo, Bucci, Magri, Magri, & Marini, 2014; Borràs et al., 2015). Thus, the fusion of data from complementary techniques can provide more accurate knowledge in metabolomics studies. This merged information could improve the results, reducing the error rate in comparison with those obtained by a single technique (Callao & Ruisánchez, 2018).

Herein, we present the first evaluation of the NMR based metabolic profiles for golden rums classification based on diverse factors, such as barrel used for maturation, raw material selection and aging. Different golden rums were analysed by ¹H NMR experiments with water and ethanol suppression followed by multivariate analysis. PCA models were built to highlight possible differences among the overall structure of the metabolite data, identifying also potential outliers. Different multivariate analyses were applied to build discriminant models. Additionally, and for the first time in food matrices, data fusion strategies with GC-MS and LC-MS data previously reported were applied including volatile, non-volatile and major components evaluation.

2. Experimental section

2.1. Materials and reagents

Deuterium oxide (D₂O, D ≥ 99.9%) and Norell™ 5.0 mm O.D. × 178 mm borosilicate NMR sample tubes and sodium-3-trimethylsilylpropionate (TMSP-2,2,3,3-d₄, 98%) were purchased from Euroisotop (St-Aubin Cedex, France). Milli-Q water (Millipore, Bedford, MA, USA) was used for method optimization and calibration.

2.2. Apparatus

NMR spectra were obtained on a Bruker AVANCE III HD 600 MHz spectrometer equipped with a QCI ¹H/¹³C/¹⁵N/³¹P proton-optimized quadrupole inverse cryoprobe with ¹H and ¹³C cryochannels. Chemical shifts (δ) are given in parts per million (ppm). ¹H chemical shifts are given in comparison to TMSP-2,2,3,3-d₄ (¹H = 0.0 ppm).

2.3. Samples and sample preparation

Twenty-four commercial rums from 10 different countries were purchased in several supermarkets (Almería, Spain). All samples were stored in their original glass bottles prior analysis at room temperature. Table 1 shows rum information obtained from the label and the official website of manufacturers, and different codes were assigned for each rum in this study. For those rums whose information was not available, they were described as N/A. Those rums that did not provide enough information about their elaboration have been analyzed as unknown (N/A) samples.

A simple pre-treatment method was carried out in rum samples: 0.5 mL of the rums were directly analysed after addition of 100 µL of deuterated water and 0.05% (w/v) internal standard reference (TMSP). The resulting solution was placed into a 5 mm NMR tube and directly analysed in an NMR Bruker AVANCE 600 MHz spectrometer.

2.4. ¹H NMR method

¹H NMR spectra were recorded at an operating frequency of 600.130 MHz and a sweep width of 12019.230 kHz (20.03 ppm). A Bruker sequence ZG30 was used to calibrate the suppression of water and ethanol signals. The water signal was suppressed at 4.79 ppm with an integration of 310.73 Hz (155.36 Hz on both sides of the signal centre). Each ethanol signal was suppressed with an integration of 145.45 Hz (72.73 Hz on each side of the signal centre). Additionally, rum analyses were made decoupling on the ¹³C channel using a garp4 pulse decoupling sequence during acquisition. These integrations were fixed to improve baseline and avoid deformations due to excessive suppression. Thirty-two free induction delays (FIDs) were collected. Finally, one dimensional ¹H NMR pulse using presaturation during relaxation delay and mixing time using shaped pulse for off-resonance presaturation to suppress the water and ethanol signals was used (LC1PNGPPS). Shimming was performed on each sample prior to data acquisition using the TopShim automatic shimming method from Bruker BioSpin software. All samples were measured at 27.0 ± 0.1 °C and under nonspinning conditions with regard to the high magnetic field to avoid spinning side bands. NMR spectra were collected using the following acquisition parameters: size of fid (TD) = 65 K, acquisition time (AQ) = 2.7 s, relaxation delay (D1) = 2.4 s, and an FID resolution (FIDRES) = 0.37 Hz. The Lorentzian line-broadening factor of 1 Hz was applied to the total FID prior to Fourier transformation. The spectra were automatically phased, baseline-corrected and integrated using TOPSPIN software (version 4.0). The spectrometer transmitter was locked to D₂O frequency throughout each analysis, and the receiver gain value was calculated automatically.

Table 1
Code, bottle labelling and manufacturer information for the rum samples.

Brand code	Bottle code	Origin	Aging	Raw material	Barrel	Distillation method
TAB	TAB	Caribbean	Young	N/A	N/A	N/A
HAV	HAV_5 HAV_7	Cuba	5 years 7 years	Molasses Molasses	French oak barrel French oak barrel	Stainless steel column Stainless steel column
LEG	LEG_A LEG_B		Anejo Young	Molasses Molasses	American oak barrel No barrel	N/A N/A
MAT	MAT_7 MAT_10 MAT_15	Dominican Republic	7 years 10 years 15 years	Molasses Molasses Molasses	American oak barrel American oak barrel American oak barrel	Stainless steel column Stainless steel column Stainless steel column
ALM	ALM		Anejo	N/A	American oak barrel	N/A
BAR	BAR		Old reserve	N/A	N/A	N/A
BOH	BOH		Anejo	N/A	N/A	N/A
BOT	BOT_8 BOT_12 BOT_18	Guatemala	8 years 12 years 18 years	Sugarcane juice Sugarcane juice Sugarcane juice	American oak barrel American oak barrel American oak barrel	Stainless steel column Stainless steel column Stainless steel column
APP	APP_5 APP_12	Jamaica	5–10 years 12 years	Molasses Molasses	American oak barrel American oak barrel	Copper pot steel Copper pot steel
FLOR	FLOR_7 FLOR_12 FLOR_18	Nicaragua	7 years 12 years 18 years	Molasses Molasses Molasses	American oak barrel American oak barrel American oak barrel	Stainless steel column Stainless steel column Stainless steel column
NEG	NEG	Spain	Anejo	Molasses	N/A	N/A
REY	REY		N/A	N/A	N/A	N/A
ST	ST_3 ST_10	Venezuela	3 years 10 years	Molasses Molasses	French oak barrel French oak barrel	Stainless steel column Stainless steel column
VEL	VEL		Young	N/A	N/A	N/A

2.5. Pre-processing and data treatment

The spectral data were imported into NMRProcFlow 1.2 (Jacob, Deborde, Lefebvre, Maucourt, & Moing, 2017), and alignment was performed, water and ethanol regions were removed, and unitary bucketing (0.1 ppm) was performed. One hundred buckets were designed for the obtained NMR data sets. Data set was imported as an XML and then processed. Excel software, version 2013 (Redmond, WA, USA) and SIMCA software (Umetrics, Umeå, Sweden) (Triba et al., 2015) were used, respectively.

2.6. Chemometric methods

For exploratory analysis and visual interpretation of the analytical connection among the samples, unsupervised analyses, such as PCA, were applied. Using Hotelling's T^2 analysis, potentially variables disturbing the proposed models (outliers) were detected. On the other hand, DModX graphic was applied to propose additional moderate outliers (Marini, 2010). Detected outliers were excluded for further statistical analyses.

Rums were modelled in relation to the settled classes (Y) using supervised PLS-DA. The variable importance in projection (VIP) analysis was used for the interpretation of the discrimination ability of the variables in the proposed model of rums.

The coefficient of determination (R^2X), the proportion of the variance of the response that is explained by the model (R^2Y) and the predictive ability parameter (Q^2Y) allowed the correct evaluation of the models in terms of goodness-of-fit. The k-fold cross-validation (k-fold-CV) allowed the ability of the model estimation, avoiding possible overfitting. With this aim, the number of significant components (NSCs) were estimated (Triba et al., 2015). Additionally, receiver operating characteristic (ROC) curve analysis was applied for the validation of binary and triple classifications (Broadhurst & Kell, 2006).

2.7. Multi-technique data fusion

In order to obtain the maximum number of correctly classified

samples, data fusion of different instrumental techniques was applied; 1H NMR, LC-MS and GC-MS. As each technique is scale dependent, adequate preprocessing of the obtained data was necessary in order to avoid systematic variations and obtaining adequate datasets.

For that purpose, low-level and mid-level fusion strategies were applied (also known as variable-data fusion). Thus, rum analysis data previously obtained by LC coupled with high-resolution MS (Exactive-Orbitrap analyser) (Belmonte-Sánchez et al., 2019) and headspace solid phase microextraction (HS-SPME) coupled to GC-MS (Belmonte-Sánchez et al., 2018) were applied.

In these previous studies, HS-SPME and "dilute and shoot" (using Milli-Q water as extraction solvent) extraction were used as generic sample treatment in order to extract as many compounds as possible. In this way, the comparison of the results in the data fusion can be considered as a data set with minimum loss of information due to the extraction step.

For the low-level fusion, classification results were obtained unifying all the data into a unique matrix containing all the information from the different instruments before building the model (Biancolillo et al., 2014). The variable data-fusion was also tried, extracting relevant features from the different techniques and then, concatenating them into a single array for further multivariate classification. The selection of the most discriminant compounds, based on discriminant ability value for each model, was applied for the mid-level fusion (Obisesan, Jiménez-Carvelo, Cuadros-Rodríguez, Ruisánchez, & Callao, 2017).

2.8. Model validation-prediction ability

The reliability of all supervised analyses was ensured and for that purpose all the samples were randomly divided in two differentiated blocks: 80% of the samples were selected for the statistical modelling, meanwhile the remaining 20% were employed as validation set. The percentage of samples correctly classified into their respective groups (% CC) was calculated to evaluate the established models in an external way. To determine the accuracy of prediction, root-mean-square error of cross-validation (RMSECV) was calculated (Xi, Gu, Baniyasi, & Raftery, 2014). Measure of significance for the observed group

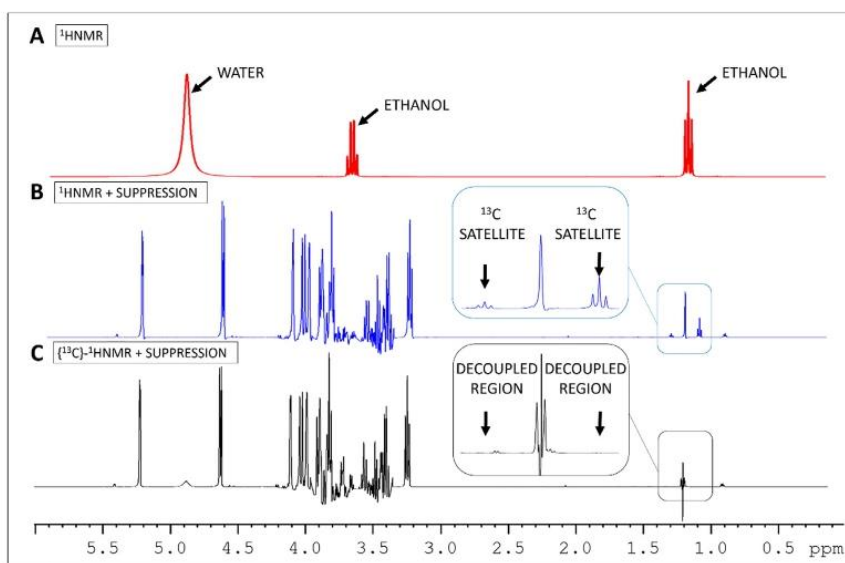


Fig. 1. Spectra between 6.0 and 0.0 ppm 600 MHz ^1H NMR: a) full spectrum obtained for BOT 12 rum, b) ethanol and water suppressed spectrum and c) one bond ^{13}C satellite-suppressed ^1H spectrum acquired.

separation by the obtained variables (p -value) was carried out by cross-validated ANOVA analysis (CV-ANOVA) (Eriksson, Trygg, & Wold, 2008). To avoid overfitting of the proposed discriminant analysis methods, DA-validation was carried out by a permutation test, and intercept value of $R^2 < 0.3$ and $Q^2 < 0.05$ indicated a statistical significance and no overfitting with high predictive value of the model (Liu et al., 2018).

3. Results and discussion

3.1. ^1H NMR acquisition

TMSP was used as internal standard for normalization and alignment of the rum spectra because it is a water-soluble compound compatible with ^1H NMR experiments, giving a typical chemical shift at δ_{H} 0.00 ppm, where rum samples do not present any proton signal.

Rums contain major components such as water, ethanol and sugars, whose NMR signals overlap with those from other matrix components. The signals for ethanol (1.17 and 3.65 ppm), residual water (4.76 ppm), and TMSP chemical shift reference (0.00 ppm) are shown in Fig. 1A. To obtain the maximum quality of the spectral information, ethanol and water suppression were applied; the appearance of numerous signals related to the minority compounds of the rums among the spectra allowed further statistical analysis based on compositional differences (Fig. 1B). Moreover, the possibility of eliminating the ^{13}C -satellites of the suppressed signals was exploited using a pulse decoupling sequence during acquisition (Fig. 1C).

As can be seen in Fig. 1, the appearance of proton signals in the 3.0–5.5 ppm region was very clear. This signal overlapping is directly correlated with the high concentration of carbohydrates (sugars) of alcoholic beverages typically found by NMR, such as α -D-fructofuranose/ β -D-fructofuranose (4.0–4.2 ppm approximately), α -D-glucopyranose (5.2–5.3 ppm approximately) and sucrose (4.3–5.5 ppm approximately) (Hohmann et al., 2015; Sánchez-Estébanez et al., 2018).

3.2. Data pre-processing

All the rums were processed in triplicate and imported into NMRProcFlow 1.2 software for alignment (chemical shift calibration to

TMSP internal standard, 0.0 ppm). Each spectral region of interest or bucket was determined with the unitary bucketing module for all the samples batch. Buckets of 0.1 ppm of width (S/N 0 threshold) were selected within 0.0–10.0 ppm, dividing the region into one hundred sequential segments. Bucket of 0.1 ppm broadening was selected in order to cover typical spectra widening for detected signals among rum spectra. The spectral region containing the suppressed water and ethanol signals was excluded from the bucketing process, improving further statistical results. The resulting data set was exported as an Excel XML file before chemometric analyses by SIMCA software.

For PCA and PLS-DA models, pareto scaling ($1/\sqrt{\text{SD}}$) is applied to increase the discriminant potential of those minor signals that may be affected by the presence of majority analytes (Worley & Powers, 2012). Additionally, logarithmic transformation was used for all the dataset, obtaining values in a smaller range without masking the effect of small values in the dataset and allowing an optimal variable range for the proposed model (Lubes & Goodarzi, 2017).

3.3. Chemometric methods

First, both ^{13}C coupled and decoupled ^1H NMR results were compared in terms of classification performance of the unsupervised and supervised models ($R^2\text{X}$, $R^2\text{Y}$ and $Q^2\text{Y}$). Although good results were obtained for the $^1\text{H}(^{13}\text{C})$ -NMR models, better results were obtained for ^1H NMR models (see Table S-1). These results can be explained in terms of spectral resolution degradation of the ^1H NMR observed signals due to decoupling schemes, which can degenerate the compromise between sensitivity and experimental resolution (Moutzouri et al., 2017).

PCA unsupervised analysis was performed within all the rums, with the aim of showing intergroup separation trending on the scores plot (Fig. 2). There were no clear groups due to the fact that metabolomic differences are usually minor in comparison to intra- and inter-observation variations. Nevertheless, the analysis of the obtained model plot (two principal components) allowed a visual trend considering rum aging, as well as excellent reproducibility of the replicate analyses. Linear trend among different aged rums from different brands can be observed in Fig. 2 (arrows were added indicating aging trend within rums from the same brand). Furthermore, the wide dispersion of the observations among the model denoted excellent heterogeneity of the

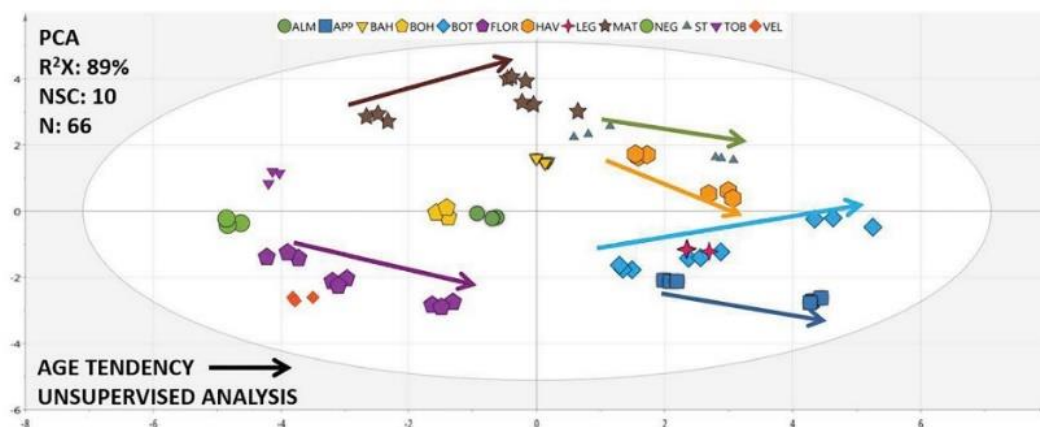


Fig. 2. PCA representation of the remaining 66 rums after data pre-processing and 89% of variance (NSC = 10).

present metabolomic study.

For further classification of the selected aging, raw material, barrel, distillation method and origin groups, different PLS-DA models were built. Initial identification based on the age of the rums was carried out by considering samples aged equal to or < 7 years as young rums, and samples older than 7 years, as old rums. Additionally, in order to validate the model in terms of recognition ability (% CC), 13 observations were assigned for the validation set.

For PLS-DA analyses, significant models were obtained. Fisher's probability $p < 0.05$ (95% confidence) was also calculated discarding the probability of the classification occurring by chance. All the additional evaluation values of the proposed models are shown in Table S-2, showing appropriate results.

Initially, a PLS-DA model based on the classification of the rums by origin was carried out. As it has been demonstrated in previous rum studies, the origin of the rums is not directly correlated with its composition (Belmonte-Sánchez et al., 2018, 2019). Most of the differences between different rum brands are due to diversity in the manufacture production processes, although other factors such as sensory differences were not included in this study.

As it is shown in Table S-3, good results were obtained for PLS-DA based on aging, barrel type, distillation method and raw material. Fig. 3A shows how the classification by aging lead to a clear separation in terms of young and old rums. A model of rum classification based on young and old rums was obtained, providing an objective prediction of the age of rum despite the mixture of rums of different ages that occurs during the preparation processes. Additionally, such clear differences allowed further aging models based on OPLS-DA, obtained after the elimination of variables that are not correlated with the main components selected (Trygg & Wold, 2002). This approach improves the prediction ability of the age of the rums ($R^2Y = 0.974$; $Q^2Y = 0.913$) (Fig. S-1), without compromising the model fitting (permutation test results; $R^2 = 0.85$, $Q^2 = -0.63$).

For the PLS-DA models based on barrel type, distillation method and raw material, a better understanding was achieved for those rums when no information was available. The classification ability of models showed a 95.5–100% correct prediction of the validation test samples (CC%), allowing for classifying rums in the proposed categories with excellent reliability (Table S-2).

Due to graphical proximity in the proposed supervised models, differences in the two first components bi-plots can be seen for those rums whose manufactures do not specify how rums were made. When the type of barrel used is not described, wood similarities have been found within the American oak wood (Fig. 3B). This behaviour can be explained due to variations in the controlled oxidation and

modification of tannins from the wood, and anthocyanins from previous alcoholic beverages aged in the barrels, such as wine (Kyralcou et al., 2016). Improvement of the flavour, taste and stability of the rums is then achieved. With this aim, additional treatments as oak chips or micro-oxygenation can be applied for obtaining optimal maturation conditions, increasing the production cost (Crump, Johnson, Wilkinson, & Bastian, 2015).

Clear evidences of differentiation in the distillation process (Fig. 3C) and raw material (Fig. 3D) were found as well, which can be attributed to different conditions in the distillation process (temperature, duration, fraction separation procedure, etc.) and quality of the sugary materials, respectively. Even though copper steel distillations have demonstrated the positive effect in the reduction of sulphur compounds in spirit beverages, which are responsible for unpleasant aromas, (Harrison, Fagnen, Jack, & Brosnan, 2011), higher costs are attributed to this technique due to that further reduction of copper concentration (usually by cationic exchange resins) is needed (Nóbrega, Pereira, Paiva, & Lachenmeier, 2011). Minimisation of undesirable metals can be reached by stainless steel distillations, getting in addition a considerable reduction in the final production price. Taking into account this fact, unspecified rums group can be clearly correlated to stainless steel distillations (Fig. 3C).

The differences showed in Fig. 3D can be explained taking into account that molasses is the principal by-product of the refining of sugarcane raw material, which is determinant for the differentiation of rums (Medeiros, de Matos, de Pinho Monteiro, de Carvalho, & Soccol, 2017). The PLS-DA two-dimensional representation allowed for totally discarding the application of molasses in the manufacturing process of those unspecified rums. It should be noted that one of the analysed rums (Fig. 3D) showed an outstanding distance to the proposed model, indicating that the sugary material used for fermentation is not common in conventional rum preparation process. This rum was discarded for following chemometric analyses.

3.4. Multi-technique data fusion

The possibility of combining information from different instruments by means of low- and mid-level data fusion strategies was also investigated. Both data fusion levels were applied in order to improve the obtained individual models (GC-MS, LC-MS and 1H NMR). With this aim, the data from GC-MS and LC-MS techniques were extracted from previous rum metabolomics studies carried out by our research team (Belmonte-Sánchez et al., 2018, 2019). Due to variations in the sampling approach in different studies, not all rum samples were included in the three studies, and finally a total of 64 observations and 1352

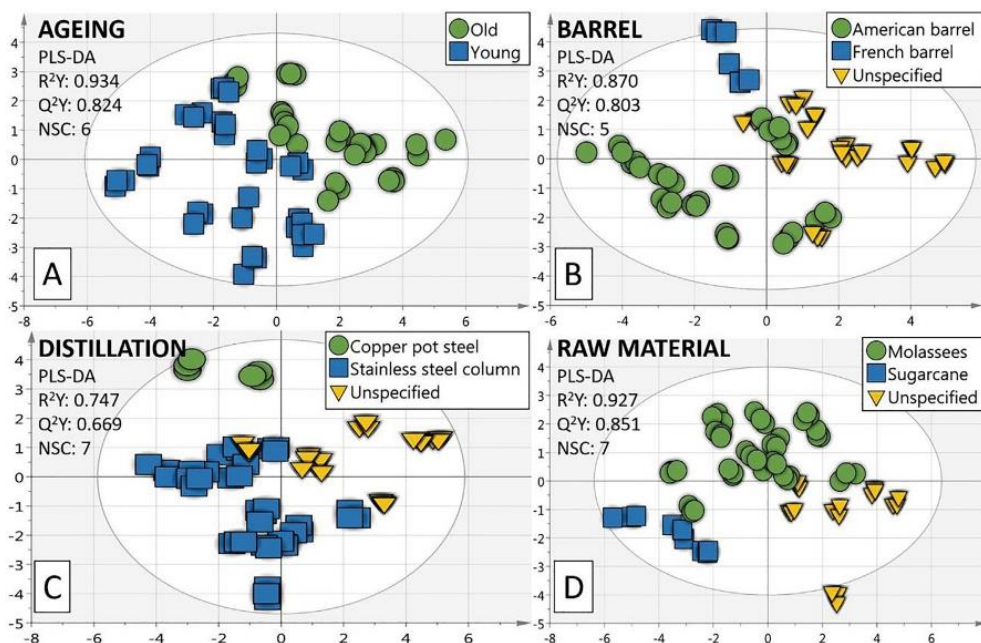


Fig. 3. PLS-DA representation showing the classification of the rums batch on the basis of: (A) aging, (B) barrel type, (C) distillation method and (D) raw material.

Table 2

Validation values obtained for each classification model of the rums by low-data fusion method.

Model class	Aging	Raw material	Barrel type	Distillation method
R ² Y	0.987	0.987	0.986	0.973
Q ² Y	0.912	0.938	0.952	0.917
NSC	5	6	7	6
Permutation test*	0.86 (-0.41)	0.67 (-0.73)	0.84 (-0.70)	0.70 (-0.70)
p-Value	< 0.01	< 0.01	< 0.01	< 0.01
Fishefs probability	1.6e-14	1.3e-19	4.9e-11	3.4e-14
ROC**	0.99 (Old) 0.99 (Young)	0.93 (Unspecified) 0.98 (Molasses) 1.00 (Sugarcane)	1.00 (Unspecified) 0.99 (American) 1.00 (French)	1.00 (Unspecified) 1.00 (Copper pot) 1.00 (Stainless steel)
CC%	100.0	100.0	100.0	100.0

*Permutation test results are expressed as “R2 (Q2)” values.

** ROC results are expressed as Area Under the Curve (AUC).

variables (100 variables for NMR, 232 variables for GC-MS and 1020 variables for LC-MS) were processed and collected in the resulting XML file. The same data process described in Section 2 was applied for the multi-technique data fusion results, showing the obtained results in Table 2.

As it can be seen in Table 2 the low-level data fusion strategy provided significantly better classification results than the individual techniques (Table S-3). R²Y higher than 0.97 for all the four classes and a successfully 100% classification value of CC% for all the test set samples were obtained, improving the statistical parameters of the model. As can be seen in Fig. 4A, a new separation can be discerned, corresponding to middle aged rum from 7 to 8 years. For barrel type (Fig. 4B), French barrel was totally discarded for non-labelled rums. The best results were obtained for the classification of rums depending on both distillation method and raw material (Fig. 4C and D). Clear evidences of differences in the production (e.g. distillation conditions) and possible lower quality sugar fermentation were found.

Due to the three data blocks used in the building of single model on fused data, different contributions of the individual blocks are obtained (Table S-3). Assuming all the obtained results as adequate for

metabolomics studies (obtained R²Y from 0.75 to 0.98 and Q²Y from 0.67 to 0.92), ¹H NMR has proven to be the most suitable technique for aging classification (Table S-3). For raw material, barrel type and distillation method classification, LC-MS provided the most acceptable results, assuming that considerable number of variables (1020 chromatographic peaks) were used. Therefore more sophisticated and time consuming data treatment is needed. On the other hand, GC-MS showed an excellent behaviour in the four selected categories, ensuring reliable and validated results within 232 variables, obtaining R²Y from 0.92 to 0.97 and Q²Y from 0.71 to 0.84.

Additionally, middle-fusion level was evaluated, based on the selection of the most discriminant compounds for each different technique model. With this aim, VIP analysis was applied for variable selection; VIP > 1.5 for LC-MS and VIP > 1.0 for GC-MS and ¹H NMR for each classification category. Larger VIP value was selected for LC-MS due the higher number of variables (1020 variables), which leads to the proposed VIP value in order to obtain reasonable data reduction. By this way, variable data volume was considerably reduced to 50 variables, simplifying data treatment and processing time. In the case of middle-fusion level, although excellent results were achieved for

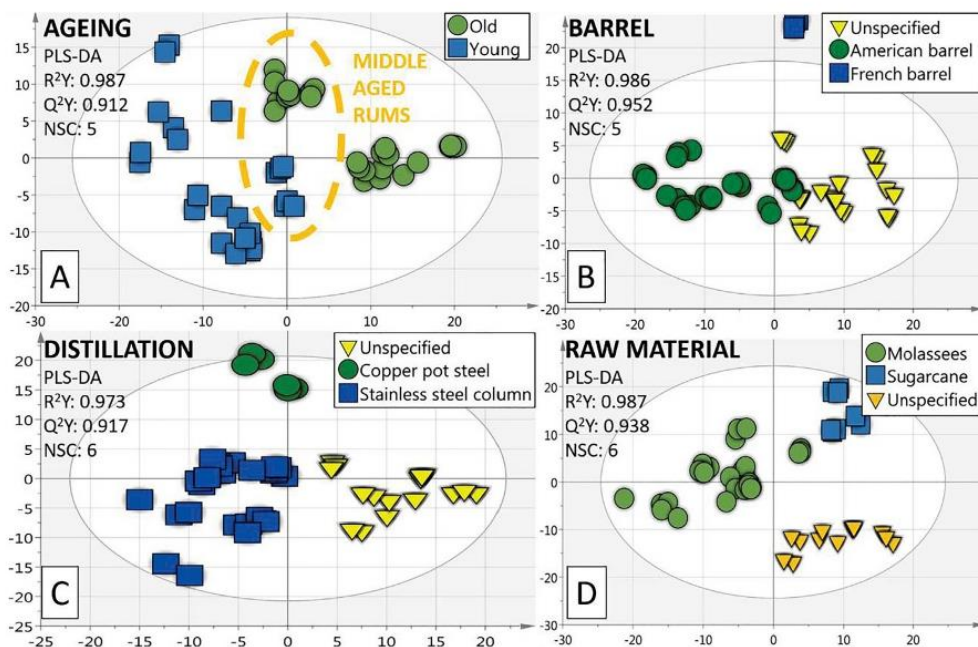


Fig. 4. PLS-DA representation showing the classification of the rums batch using low-data fusion method on the basis of (A) aging, (B) barrel type, (C) distillation method and (D) raw material.

all the proposed categories, obtaining R^2Y from 0.90 to 0.88 and Q^2Y from 0.85 to 0.78 (see Table S-4), it is evident the possible influence of the information loss due to significant data reduction.

4. Conclusions

In this study, the potentiality of an analytical 1H NMR strategy combined with chemometrics for the classification of golden rums depending on different factors was demonstrated. Significant results were obtained for all the supervised PLS-DA models, showing excellent fitting (R^2Y from 0.93 to 0.75) and predictive abilities (Q^2Y from 0.85 to 0.67) for the selected categories: type of barrel, raw material, distillation method and aging. For those samples whose label did not provide specific information about the proposed categories, the obtained models provided differences in the quality of the used materials and distillation conditions. For aging classification, despite the complexity of the rums blending process, the model obtained using 1H NMR data correctly classified the rums by old and young ones.

Lastly, the potential of a multiplatform instrumental fingerprinting using NMR, GC-MS and LC-MS by low-level data fusion has proved to provide the best results in comparison with the individual techniques (100% of correctly classified samples for all proposed categories). Those results can be applied in different scenarios with different perspectives, as 1H NMR technique has showed to be a very fast (no pre-treatment of sample is needed) way to obtain a very suitable metabolomic model. Therefore more accurate models can be obtained in relation to GC-MS and LC-MS, where larger analysis and more sophisticated data treatment are needed.

CRedit authorship contribution statement

José Raúl Belmonte-Sánchez: Formal analysis, Investigation, Writing - original draft, Visualization. **Roberto Romero-González:** Methodology, Software, Data curation. **José Luis Martínez Vidal:** Conceptualization, Funding acquisition. **Francisco Javier Arrebola:**

Validation. **Antonia Garrido Frenich:** Resources, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2020.126363>.

References

- Belmonte-Sánchez, J. R., Gherghel, S., Arrebola-Liébanas, J., Romero González, R., Martínez Vidal, J. L., Parkin, I., & Garrido Frenich, A. (2018). Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase micro-extraction coupled to gas chromatography-mass spectrometry. *Talanta*, 187(May), 348–356. <https://doi.org/10.1016/j.talanta.2018.05.025>.
- Belmonte-Sánchez, J. R., Romero-González, R., Arrebola, F. J., Vidal, J. L. M., & Garrido Frenich, A. (2019). An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography-orbitrap mass spectrometry and chemometric strategies. *Journal of Agricultural and Food Chemistry*, 67(4), 1302–1311. <https://doi.org/10.1021/acs.jafc.8b05622>.
- Bharti, S. K., & Roy, R. (2012). Quantitative 1H NMR spectroscopy. *TrAC – Trends in Analytical Chemistry*, 35(May), 5–26. <https://doi.org/10.1016/j.trac.2012.02.007>.
- Biancolillo, A., Bucci, R., Magri, A. L., Magri, A. D., & Marini, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Analytica Chimica Acta*, 820, 23–31.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment – A review. *Analytica Chimica Acta*, 891, 1–14. <https://doi.org/10.1016/j.aca.2015.04.042>.
- Broadhurst, D. I., & Kell, D. B. (2006). Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, 2(4), 171–196. <https://doi.org/10.1007/s11306-006-0037-z>.
- Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control*, 86, 283–293. <https://doi.org/10.1016/j.foodcont.2017.11.034>.
- Coldea, T. E., Mudura, E., & Socaciu, C. (2017). Advances in distilled beverages

- authenticity and quality testing. *Ideas and Applications Toward Sample Preparation for Food and Beverage Analysis*. <https://doi.org/10.5772/intechopen.72041>.
- Crump, A. M., Johnson, T. E., Wilkinson, K. L., & Bastian, S. E. P. (2015). Influence of oak maturation regimen on composition, sensory properties, quality, and consumer acceptability of cabernet sauvignon wines. *Journal of Agricultural and Food Chemistry*, 63(5), 1593–1600. <https://doi.org/10.1021/jf5044025>.
- da Silva, L. A., Flumignan, D. L., Tininis, A. G., Pezza, H. R., & Pezza, L. (2019). Discrimination of Brazilian lager beer by ¹H NMR spectroscopy combined with chemometrics. *Food Chemistry*, 272(August 2018), 488–493. <https://doi.org/10.1016/j.foodchem.2018.08.077>.
- Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for significance testing of PLS and OPLS* models. *Journal of Chemometrics*, 22(11–12), 594–600. <https://doi.org/10.1002/cem.1187>.
- European Commission (2008). EC Regulation 110/2008 of the European parliament and of the council on the definition, description, presentation, labelling and the protection of geographical indications of spirit drinks. *Official Journal of the European Union*, L39, 16–54.
- Farag, M. A., Khattab, A. R., Ehrlich, A., Kropf, M., Heiss, A. G., & Wessjohann, L. A. (2018). Gas chromatography/mass spectrometry-based metabolite profiling of nutrients and antinutrients in eight lentil and lupinus seeds (fabaceae). *Journal of Agricultural and Food Chemistry*, 66(16), 4267–4280. <https://doi.org/10.1021/acs.jafc.8b00369>.
- Fortunato, W., Rocha, D. C., & Sheen, D. A. (2018). Classification of samples from NMR-based metabolomics using principal components analysis and partial least squares with uncertainty estimation. *Analytical and Bioanalytical Chemistry*, 6305–6319. <https://doi.org/10.1007/s00216-018-1240-2>.
- Fotakis, C., & Zervou, M. (2016). NMR metabolic fingerprinting and chemometrics driven authentication of Greek grape marc spirits. *Food Chemistry*, 196, 760–768. <https://doi.org/10.1016/j.foodchem.2015.10.002>.
- Gougeon, L., da Costa, G., Le Mao, I., Ma, W., Teissedre, P. L., Guyon, F., & Richard, T. (2018). Wine analysis and authenticity using 1H-NMR metabolomics data: Application to Chinese wines. *Food Analytical Methods*, 2018, 1–10. <https://doi.org/10.1007/s12161-018-1310-2>.
- Harrison, B., Fagnen, O., Jack, F., & Brosnan, J. (2011). The impact of copper in different parts of malt whisky pot stills on new make spirit composition and aroma. *Journal of the Institute of Brewing*, 117(1), 106–112. <https://doi.org/10.1002/j.2050-0416.2011.tb00450.x>.
- Hohmann, M., Koospl, V., Bauer-Christoph, C., Christoph, N., Wachter, H., Diehl, B., & Holzgrabe, U. (2015). Quantitative ¹H NMR analysis of egg yolk, alcohol, and total sugar content in egg liqueurs. *Journal of Agricultural and Food Chemistry*, 63(16), 4112–4119. <https://doi.org/10.1021/acs.jafc.5b00940>.
- Hoyos Ossa, D. E., Gil-Solsona, R., Peñafla, G. A., Sancho, J. V., & Hernández, F. J. (2018). Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics. *Food Chemistry*, 250, 89–97. <https://doi.org/10.1016/j.foodchem.2018.01.038>.
- Jacob, D., Deborde, C., Lefebvre, M., Maucourt, M., & Moing, A. (2017). NMRProcFlow: A graphical and interactive tool dedicated to 1D spectra processing for NMR-based metabolomics. *Metabolomics*, 13(4), 1–5. <https://doi.org/10.1007/s11306-017-1178-y>.
- Kyraleou, M., Tzanakouli, E., Kotseridis, Y., Chira, K., Ligas, I., Kallithraka, S., & Teissedre, P.-L. (2016). Addition of wood chips in red wine during and after alcoholic fermentation: Differences in color parameters, phenolic content and volatile composition. *OENO One*, 50(4), 209–222. <https://doi.org/10.20870/oeno-one.2016.50.4.885>.
- Lea, A. G. H., & Piggott, J. (2012). *Fermented beverage production*. Springer Science & Business Media.
- Liu, J., Chen, N., Yang, J., Yang, B., Ouyang, Z., Wu, C., ... Chen, M. (2018). An integrated approach combining HPLC, GC/MS, NIRS, and chemometrics for the geographical discrimination and commercial categorization of saffron. *Food Chemistry*, 253, 284–292. <https://doi.org/10.1016/j.foodchem.2018.01.140>.
- Longobardi, F., Innamorato, V., Di Gioia, A., Ventrella, A., Lippolis, V., Logrieco, A. F., ... Agostiano, A. (2017). Geographical origin discrimination of lentils (*Lens culinaris* Medik.) using ¹H NMR fingerprinting and multivariate statistical analyses. *Food Chemistry*, 237, 743–748. <https://doi.org/10.1016/j.foodchem.2017.05.159>.
- Lubes, G., & Goodarzi, M. (2017). Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chemical Reviews*, 117(9), 6399–6422. <https://doi.org/10.1021/acs.chemrev.6b00698>.
- Marini, F. (2010). Classification methods in chemometrics. *Current Analytical Chemistry*, 6(1), 72–79. <https://doi.org/10.2174/157341110790069592>.
- Medeiros, A. B. P., de Matos, M. E., de Pinho Monteiro, A., de Carvalho, J. C., & Soccol, C. R. (2017). Cachaça and rum. *Current Developments in Biotechnology and Bioengineering* (pp. 451–468). Elsevier <https://doi.org/10.1016/B978-0-444-63666-9.00016-9>.
- Monakhova, Y. B., Schäfer, H., Humpfer, E., Spraul, M., Kuballa, T., & Lachenmeier, D. W. (2011). Application of automated eightfold suppression of water and ethanol signals in ¹H NMR to provide sensitivity for analyzing alcoholic beverages. *Magnetic Resonance in Chemistry*, 49(11), 734–739. <https://doi.org/10.1002/mrc.2823>.
- Moutzouri, P., Kiraly, P., Phillips, A. R., Coombes, S. R., Nilsson, M., & Morris, G. A. (2017). 13C satellite-free 1H NMR spectra. *Analytical Chemistry*, 89(22), 11898–11901. <https://doi.org/10.1021/acs.analchem.7b03787>.
- Mung, D., & Li, L. (2018). Applying quantitative metabolomics based on chemical isotope labeling LC-MS for detecting potential milk adulterant in human milk. *Analytica Chimica Acta*, 1001, 78–85. <https://doi.org/10.1016/j.aca.2017.11.019>.
- Navarro, M., Kontoudakis, N., Gómez-Alonso, S., García-Romero, E., Canals, J. M., Hermosín-Gutiérrez, I., & Zamora, F. (2018). Influence of the volatile substances released by oak barrels into a Cabernet Sauvignon red wine and a discolored Macabeo white wine on sensory appreciation by a trained panel. *European Food Research and Technology*, 244(2), 245–258. <https://doi.org/10.1007/s00217-017-2951-x>.
- Nóbrega, I. C. C., Pereira, J. A. P., Paiva, J. E., & Lachenmeier, D. W. (2011). Ethyl carbamate in cachaça (Brazilian sugarcane spirit): Extended survey confirms simple mitigation approaches in pot still distillation. *Food Chemistry*, 127(3), 1243–1247. <https://doi.org/10.1016/j.foodchem.2011.01.125>.
- Obisesan, K. A., Jiménez-Carvelo, A. M., Cuadros-Rodríguez, L., Ruisánchez, I., & Callao, M. P. (2017). HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil. *Talanta*, 170, 413–418. <https://doi.org/10.1016/j.talanta.2017.04.035>.
- Ralli, E., Amargianitaki, M., Manolopoulou, E., Misiak, M., Markakis, G., Tachtalidou, S., & Spyros, A. (2018). NMR Spectroscopy Protocols for Food Metabolomics Applications. *Metabolic Profiling Methods and Protocols. Methods in Molecular Biology*, 1738, 203–211. https://doi.org/10.1007/978-1-4939-7643-0_14.
- Razali, M. T. A., Zainal, Z. A., Maulidiani, M., Shaari, K., Zambri, Z., Idrus, M. Z. M., & Ismail, I. S. (2018). Classification of Raw Stingless Bee Honeys by Bee Species Origins using NMR-LCMS-based Metabolomics Approach. *Molecules*, 23(9), 2160. <https://doi.org/10.20944/PREPRINTS201807.0291.V1>.
- Sánchez-Estébanez, C., Ferrero, S., Alvarez, C. M., Villafañe, F., Caballero, I., & Blanco, C. A. (2018). Nuclear magnetic resonance methodology for the analysis of regular and non-alcoholic lager beers. *Food Analytical Methods*, 11(1), 11–22. <https://doi.org/10.1007/s12161-017-0953-8>.
- Triba, M. N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., ... Savarin, P. (2015). PLS/OPLS models in metabolomics: The impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Molecular Omics*, 11(1), 13–19. <https://doi.org/10.1039/C4MB00414K>.
- Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, 16(3), 119–128. <https://doi.org/10.1002/cem.695>.
- Worley, B., & Powers, R. (2012). Multivariate analysis in metabolomics. *Current Metabolomics*, 1(1), 92–107. <https://doi.org/10.2174/2213235X11301010092>.
- Xi, B., Gu, H., Baniasadi, H., & Raftery, D. (2014). Statistical analysis and modeling of mass spectrometry-based metabolomics data. *Molecular Analysis and Genome Discovery* (pp. 333–353). Chichester, UK: John Wiley & Sons, Ltd doi: 10.1007/978-1-4939-1258-2_22.

Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabolómica de bebidas alcohólicas y al análisis de tóxicos en aguas



CAPÍTULO SEGUNDO. APLICACIÓN DE LA METABOLÓMICA
AL ESTUDIO DE RON



3 CAPÍTULO TERCERO. APLICACIÓN DE q NMR A LA DETERMINACIÓN DE TÓXICOS EN AGUAS

3.1 Introducción

Desde la aplicación de la q NMR con fines analíticos la búsqueda de nuevas aplicaciones se ha convertido en un amplio campo de estudio, permitiendo aprovechar todas las ventajas que esta técnica ofrece, más allá de su alto poder identificativo: técnica no destructiva, mínima preparación de muestra, adquisición rápida, excelente reproducibilidad y ausencia de necesidad de patrones, entre otras propiedades.

Concretamente, el estudio tanto de núcleos no dipolares (diferentes a espín $I=1/2$), como el de núcleos heteronucleares (diferentes a hidrógeno-1) supone un gran reto analítico que se presenta como una potencial alternativa para dar solución a los problemas comunes a los cuales nos enfrentamos en el laboratorio día a día.

Un nuevo caso de alto interés como analito con núcleo no dipolar (boro-11) es el del ácido bórico (H_3BO_3) (Figura 7). Se trata de un compuesto blanco, normalmente en estado de polvo, con relativa toxicidad para los seres humanos y bastante común en el medio ambiente debido a su aplicación como insecticida, herbicida y fungicida para fines alimentarios, no alimentarios y agrícolas. Puede estar presente además en forma de alguno de sus derivados naturales: tetraborato de sodio, octaborato de disodio y metaborato de sodio.

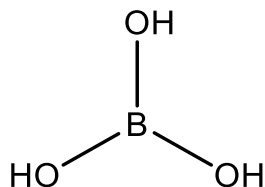


Figura 7. Estructura química del ácido bórico (H_3BO_3)

Hasta ahora, el análisis de éste tipo de analitos con núcleos de boro se ha basado en técnicas como métodos potenciométricos o ICP-MS [1,2]. La aplicación de la q NMR para la determinación de ácido bórico mediante adquisiciones ^{11}B -NMR se presenta como una excelente alternativa para el análisis de compuestos conteniendo boro, con gran potencial analítico en diversas matrices agroalimentarias, siendo la ^{11}B -NMR un modo de adquisición sin precedentes documentado en bibliografía con fines cuantitativos.

Por otro lado, un caso adicional de estudio de interés, debido a su presencia en aguas industriales por su alta capacidad anticorrosiva, es el ácido etidróico (1-hidroxi-etilideno-1,1-difosfónico, HEDP) (Figura 8). Se trata de un bifosfonato con una alta estabilidad que además puede formar complejos estables con los iones y que puede disolver los óxidos en la superficie metálica. Por ello, es un aditivo común en las plantas de procesamiento de productos agrícolas y justifica la necesidad de métodos simples y fiables para la detección de HEDP en sus aguas de lavado de productos agroalimentarios.

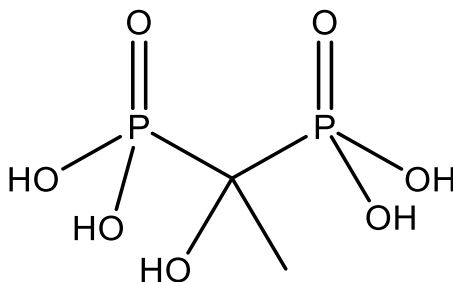


Figura 8. Estructura química del ácido etidrónico ($C_2H_8O_7P_2$)

Pese a que previamente la cuantificación del HEDP se ha llevado a cabo mediante técnicas como ICP-MS [3] o electroforesis capilar [4], su cuantificación mediante ^{31}P -NMR se presenta como una alternativa con importantes ventajas, como el tiempo de adquisición de los análisis, selectividad del fósforo gracias a su amplio rango espectral, así como su análisis directo sin pretratamiento en muestras acuosas. Además, el isótopo de fósforo-31 cuenta con un 100% de abundancia natural en la naturaleza, siendo un caso idóneo para la aplicación de la q NMR de núcleos heteronucleares en matrices agrolimentarias.

En consecuencia, el objetivo de este capítulo es el desarrollo, validación y aplicación de un método mediante NMR de boro-11 (^{11}B -NMR) para la determinación de ácido bórico en biocidas, así como mediante NMR de fósforo-31 (^{31}P -NMR) para la determinación de ácido etidrónico en agua de lavado de vegetales. Las aportaciones científicas incluidas en este capítulo son:

- **Aportación Científica 4.** *Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides*

CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA DETERMINACIÓN DE TÓXICOS EN AGUAS

- **Aportación Científica 5.** *Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative³¹P nuclear magnetic resonance method*



3.2 Bibliografía

- [1] L.-M. Zeng, H.-Y. Wang, Y.-L. Guo, Fast quantitative analysis of boric acid by gas chromatography-mass spectrometry coupled with a simple and selective derivatization reaction using triethanolamine, *J. Am. Soc. Mass Spectrom.* 21 (2010) 482–485. doi:10.1016/j.jasms.2009.12.003.
- [2] M. He, Z. Jin, H. Lu, T. Ren, Efficient separation of boron using solid-phase extraction for boron isotope analysis by MC-ICP-MS, *Anal. Methods.* 7 (2015) 10322–10327. doi:10.1039/C5AY01743B.
- [3] M. Oromí-Farrús, J.M. Minguell, N. Oromi, R. Canela-Garayoa, A reliable method for quantification of phosphonates and their impurities by ^{31}P NMR, *Anal. Lett.* 46 (2013) 1910–1921. doi:10.1080/00032719.2013.780239.
- [4] C.K. Schmidt, B. Raue, H.-J. Brauch, F. Sacher, Trace-level analysis of phosphonates in environmental waters by ion chromatography and inductively coupled plasma mass spectrometry, *Int. J. Environ. Anal. Chem.* 94 (2014) 385–398. doi:10.1080/03067319.2013.831410.

CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA DETERMINACIÓN DE TÓXICOS EN AGUAS



3.3 *Aportación científica 4. Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides*

L.M. Aguilera-Sáez, J.R. Belmonte-Sánchez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich, I. Fernández. *Analyst*. 143 (2018).

DOI: 10.1039/c8an00505b

CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA
DETERMINACIÓN DE TÓXICOS EN AGUAS



Cite this: *Analyst*, 2018, **143**, 4707

Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides†

Luis Manuel Aguilera-Sáez,  José Raúl Belmonte-Sánchez, Roberto Romero-González, José Luis Martínez Vidal, Francisco Javier Arrebola, Antonia Garrido French * and Ignacio Fernández *

Quantitative boron-11 NMR (^{11}B qNMR) spectroscopy has been introduced for the first time as a method to determine boric acid content in commercial biocides. Validation of the method affords a limit of detection of 0.02% w/w and a limit of quantification of 0.04% w/w, which are low enough to determine boric acid in commercial biocides. Other figures of merit such as linearity ($R^2 > 0.99$), recovery (93.6%–106.2%), intra- and inter-day precision (from 0.7 to 2.0%), uncertainty (3.7 to 4.4%) and matrix effects were also evaluated. This method was successfully applied to determine boric acid in five different commercial biocides in a wide range of concentrations (<0.05 to 10% w/w) providing excellent results when they were compared with those obtained using inductively coupled plasma-mass spectrometry (ICP-MS). The suitability of this method for a fast and reliable quantification of boric acid in commercial biocide preparations has been demonstrated. The absence of the matrix effect allows the application of this validated method for the determination of boric acid in other matrices of diverse composition.

Received 19th March 2018,
Accepted 26th August 2018

DOI: 10.1039/c8an00505b

rsc.li/analyst

Introduction

Nuclear magnetic resonance (NMR) has always been considered to be technically limited compared to other analytical techniques from the point of view of sensitivity. During the past decade, great improvement in electronics and materials of NMR spectrometers has been achieved. Despite these efforts, NMR is not able to reach microgram per kilogram (ppb) or nanogram per kilogram (ppt) concentration levels, which are routinely achieved by mass spectrometry.^{1,2} On the positive side, NMR provides a rich tool set for the observation of C/H/N- and X-nuclear parameters, mainly δ , J , $W_{1/2}$, and $n\text{Oe}$, allowing the construction of molecular connectivity and three-dimensional spatial relationships. Importantly, such NMR information can be obtained from isotopes of $I = 1/2$, whereas for quadrupolar nuclei the information is mainly limited to δ and $W_{1/2}$, except for those surrounded by a highly symmetric environment (low electric field gradient) or low quadrupolar moment (eQ).^{3–5}

Additionally, the quantity of an unknown compound in a mixture can be determined by comparing its relative inte-

gration area to that obtained from a known amount of a standard or using specific calibration curves.⁶ The ^1H variant of quantitative NMR (qNMR) has been the most exploited since protons are nearly ubiquitous in organic molecules. Since the first application of qNMR by Hollis in 1963,⁷ this approach has gained increasing popularity, especially in the analysis of complex natural products, such as botanical dietary supplements and traditional medicines, among many others.^{8–14}

The main aspect that makes NMR attractive for quantification is the existence of direct proportionality between the absolute integral of each signal and the number of protons giving rise to it. NMR is a non-destructive technique that simultaneously provides structural and quantitative data and allows the determination of more than one analyte in a complex mixture, even if the sample solution contains a solid phase.¹⁵

Most of the qNMR applications make use of nuclei with spin $I = 1/2$, such as ^1H , ^{19}F and ^{31}P .^{11,16–23} Quadrupolar nuclei ($I > 1/2$) have barely been introduced probably due to fundamental differences in the governing factors of the nuclear magnetic relaxation processes compared to dipolar nuclei. In fact, to the best of our knowledge, there are only a few reports where a quadrupolar nucleus has been measured for quantification purposes. Maki *et al.* have described the use of ^{27}Al qNMR ($I = 5/2$) for the quantification of various hydrolyzed species of Al^{3+} ions, such as $[\text{Al}(\text{H}_2\text{O})_6]^{3+}$, the trinuclear

Department of Chemistry and Physics, Research Centre CIAIMBITAL, Agrifood Campus of International Excellence, ceiA3, University of Almería, Ctra. Sacramento, s/n, 04120, Almería, Spain. E-mail: ifernan@ual.es

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8an00505b

complex $\text{Al}(\text{OH})_4^-$ and K-Al_{13} .¹⁷ Furthermore, there is a study that reported the use of ^{14}N qNMR ($I = 1$) for the determination of nitrate, urea and ammonium in aqueous solutions.¹⁶ All these reports provided a comprehensive overview of the often overlooked capabilities of quadrupolar qNMR. Inspired by these authors, we envisaged the possibility of extending this methodology to another quadrupolar nucleus such as boron-11. This nucleus is considered active from an NMR point of view and it is present in the structure of boric acid and its salts. It has two stable isotopes, ^{10}B and ^{11}B , with spin-quantum numbers of 3 and 3/2, respectively. ^{11}B is more suitable for NMR because of its higher sensitivity associated with its higher natural abundance (80.1 vs. 19.9%), higher gyromagnetic constant (8.58×10^7 vs. 2.87×10^7 $\text{rad s}^{-1} \text{T}^{-1}$), lower quadrupolar moment (4.06 vs. 8.46 fm^2) and better resolution as well as a higher receptivity relative to ^{13}C (777.0 vs. 23.2). Boron is also well suited to low-field instruments with a chemical shift range between +100 and -120 ppm, as has extensively been reported.^{24–30}

Traditionally, boron determinations have been achieved through the use of several analytical methods based on spectrophotometric methods,³¹ potentiometric titrations,³² acid-base titrations in the presence of auxiliary reagents, polarimetry,³³ ion chromatography with suppressed conductivity detection,³² gas chromatography coupled to MS in combination with simple derivatization reactions,³⁴ or inductively coupled plasma mass spectrometry (ICP-MS).³⁵ However, some of these methods are prone to interferences, time-consuming, and require derivatization³⁶ as well as complex analytical procedures, or several dilutions of the sample when the concentration is too high. In contrast, the method described herein based on ^{11}B qNMR has very little interferences, since nuclides other than boron are not detected, and therefore it simplifies the analysis of complex mixtures.

Boric acid-containing compounds are used as insecticides, herbicides and fungicides for non-food and agricultural purposes, among other applications.³⁷ These are available in many forms, including liquids, soluble and emulsifiable concentrates, granules, powders and pellets. In the environment, borate compounds such as borates, including boric acid, boron oxide and borate salts as sodium tetraborate, disodium octaborate and sodium metaborate are mainly found in combination with oxygen.³⁸ The concentration of each species depends on the initial concentration of boric acid and pH. Thus, at low concentrations in aqueous solutions below pH 7, boric acid exists predominantly as undissociated boric acid ($\text{B}(\text{OH})_3$), whereas at pH higher than 10, the metaborate anion $\text{B}(\text{OH})_4^-$ becomes the main species in solution. Between pH 6 and 11 and at high concentrations ($>0.025 \text{ M}$), highly water-soluble polyborate ions such as $[\text{B}_3\text{O}_3(\text{OH})_4]^-$, $[\text{B}_4\text{O}_5(\text{OH})_4]^{2-}$ and $[\text{B}_5\text{O}_6(\text{OH})_4]^-$ are formed.^{39,40}

As far as we know, ^{11}B qNMR has never been used for the quantitative analysis of boric acid-containing substances. Therefore, we report herein the development and validation of a rapid, sensitive, and reliable qNMR-based method that takes advantage of the quadrupolar nature of boron-11 towards the

accurate determination of boric acid content in commercial biocide pellets in extensive use in agrochemical applications.

Experimental section

Chemicals and materials

A total of 5 biocide pellets were acquired from different local commercial suppliers (Almería, Spain). Boric acid (H_3BO_3 , purity >99%) was purchased from Panreac (Barcelona, Spain), sodium tetraborate (Borax, purity 99%) was obtained from Sigma-Aldrich (St Louis, MO, USA) and deuterium oxide (D_2O , $D \geq 99.9\%$) and Norell™ thin-wall natural quartz NMR sample tubes (<0.1 ppm boron) were purchased from Eurisotop (St-Aubin Cedex, France). For ICP-MS analysis, Milli-Q water (Millipore, Bedford, MA, USA) was applied.

Sample preparation

All samples were ground, homogenized and stored in desiccators until NMR analysis.

One hundred fifty milligrams of homogeneous and dried biocide pellet samples were extracted in 2 mL Eppendorf® tubes with 1 mL of D_2O . The resulting mixture was vigorously vortex-stirred for 10 min and centrifuged at 5000 rpm for 10 min. Five hundred microliters of the supernatant were transferred into a previously oven-dried 5 mm quartz NMR tube. The remaining solid was washed with 1 mL of D_2O following the procedure explained previously and the supernatant was transferred into a new 5 mm quartz NMR tube. For ensuring total acid boric extraction from the remaining solid, additional extractions were performed by adding another mL of D_2O . Pellets of higher concentrations required at least three extractions in order to guarantee total recovery. For ICP-MS analysis, 500 mg of each ground sample were added to 500 mL of Milli-Q water and sonicated for 30 min, prior to analysis.

qNMR analysis

NMR spectra were recorded on a Bruker Avance 300 spectrometer (Bruker Company, Switzerland) using a direct BBFO^{PLUS} (^1H , BB- ^{19}F , SmartProbe) probe head. ^1H and ^{11}B chemical shifts are given relative to TMS and $\text{BF}_3 \cdot \text{OEt}_2$ ($^{11}\text{B} = 0$ ppm), respectively. The 90° pulse widths and attenuation levels were 9.7 $\mu\text{s}/9 \text{ W}$ for proton and 10.0 $\mu\text{s}/28 \text{ W}$ for boron.

^{11}B qNMR spectra were recorded at an operating frequency of 96.294 MHz and a sweep width of 19.231 kHz (199.7 ppm). Prior to the measurements, standard ^1H NMR spectra were always recorded in order to evaluate field homogeneity. All samples were measured at 20.0 ± 0.1 °C, without rotation and using quartz NMR tubes. NMR spectra were collected using the following acquisition parameters: number of FID data points (TD) = 64 K, acquisition time (AQ) = 0.6 s, relaxation delay (D1) = 0.1 s, FID resolution (FIDRES) = 0.58 Hz, number of scans (NS) = 64 and pre-scan delay (DE) = 6.5 μs . The Lorentzian line-broadening factor of 1.5 Hz was applied to the total FID prior to Fourier transformation. The spectra were automatically phased, baseline-corrected and integrated using



TOPSPIN software (version 3.5). Automatic integration was carried out using the multi_integ3 command. The spectrometer transmitter was locked to D₂O frequency using a mixture H₂O–D₂O (9 : 1). All NMR experiments were performed with a fixed receiver gain (RG) of 90.5, which was the optimum value estimated through several tests.

ICP-MS measurements

ICP-MS determination of boron was performed at the Research Central Services of the University of Almería. The total concentration of boron ions in dissolved commercial biocide samples was determined using an ICP-MS XSERIES 2 (Thermo Fisher, Waltham, MA, USA) equipped with an Xt interface option, a three-legged torch, a Peltier cooled (–15 °C) glass spray chamber and a PFA-20 nebulizer (operated at 0.54 L min^{–1} argon gas flow rate). The applied methodology was an adaptation of a previous method.⁴¹ The plasma was operated at 1400 W. Helium, containing 8% of hydrogen, was used as cell reagent gas and oxygen (20% oxygen in argon) was provided to the extra leg of the three-legged torch option at a flow rate of 255 mL min^{–1}.

Results and discussion

Optimization of the qNMR method

A key aspect in the development of analytical methods is the evaluation of the factors that can introduce interferences. In this sense, the use of NMR tubes made of borosilicate glass can cause interferences in the ¹¹B NMR spectra. As a result, there is a broad signal in the spectra arising from the tube and centered at δ_b –6.4 ppm that may interfere in the quantification (Fig. S1†). In order to avoid this hump, a common prac-

tice is to use quartz NMR tubes that contain low quantities of boron (<0.1 ppm), although these are more expensive and fragile than regular tubes. Consequently, quartz tubes have been used in this study. Moreover, one of the fundamental principles for quantitative analysis is that the target analyte must be completely solubilized in the selected solvent. In this study, the boric acid contained in solid biocide samples was completely extracted using D₂O.

The influence of number of scans, acquisition time and recycle delay on the accuracy of the ¹¹B qNMR method was also evaluated and the results are shown in Fig. 1. The linear correlation found within the number of scans (Fig. 1a) demonstrates that the repeatability and accuracy of the pulse generator and radio frequency transmitter, together with the linearity of the RF amplifier in the NMR equipment, are excellent. As expected, the relationship between the number of scans and the signal-to-noise ratio (SNR) follows a quadratic profile (Fig. 1b). The influence of the FID acquisition time and recycle delay on the integrated intensity is shown in Fig. 1c and d. The absolute integral of the NMR signal was almost constant for FID acquisition times or recycle delays in the range of 0.1 to 1.0 s; so, values of 0.6 s for the former and 0.1 s for the latter were set for all our measurements. The impact of these time values on the experimental time was significant since only 50 seconds of analysis were required in order to obtain a good spectrum. The T₁ relaxation time measured at a concentration of 0.5% (w/w) was 4.5 ms (Fig. S2†) which fits previously reported data.⁴² Due to the quadrupolar nature of boron-11 and the short time of analysis, ¹¹B qNMR can be proposed as a rapid and useful method able to be implemented in routine.

Since the proposed method is not based on the addition of an internal standard, a calibration curve was recorded by preparing solutions of boric acid in D₂O (without pH adjustment)

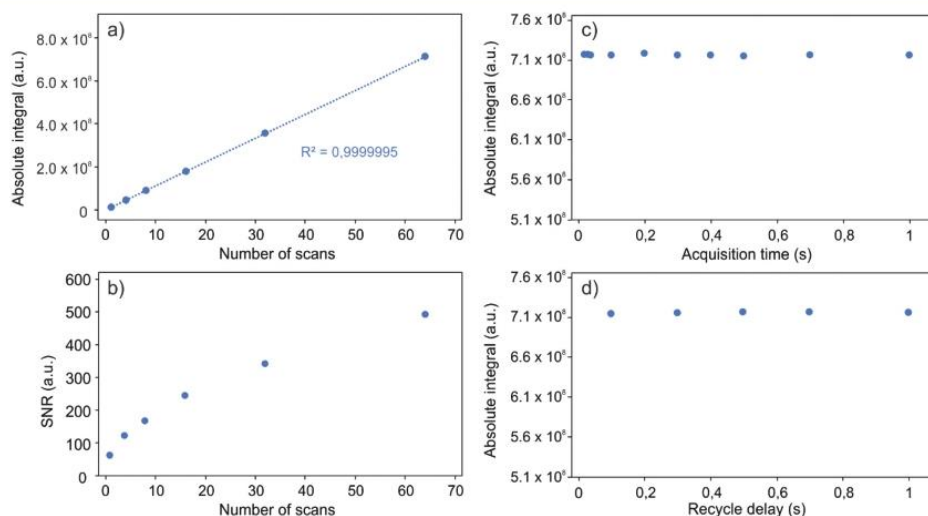


Fig. 1 Influence of: (a) number of scans on the absolute signal integral, (b) number of scans on the signal-to-noise ratio (SNR), (c) acquisition time and (d) recycle delay on the absolute signal integral.

at concentrations of 0.05, 0.10, 0.50, 1.00, 1.50, 3.00 and 4.10% (w/w). The upper concentration corresponds to the solubility limit of boric acid. It is important to mention that the pH in all these samples was 4.4, ensuring the existence of boric acid exclusively in its protonated form. Fig. 2a shows a stack of ^{11}B NMR spectra of the different boric acid solutions prepared, where each solution showed only one very sharp peak at δ_{B} 19.4 ppm. It can be observed that the boron chemical shift does not depend on the concentration of boric acid, as has been reported earlier.⁴³ Moreover, the linewidths are kept constant at all the assayed concentrations, which confirms earlier findings that show that in acidic aqueous solutions of boric acid the species $\text{B}(\text{OH})_3$ is dominant, whereas $\text{B}(\text{OH})_4^-$ can be neglected.⁴³ In order to confirm that the samples without pH adjustment are in the range where only the species $\text{B}(\text{OH})_3$ exists in solution, we reproduced the measurements on boric acid samples at 0.05, 0.10 and 1.00% (w/w) but reducing the pH down to 2.8. It was checked that the absolute integrals in the three ^{11}B qNMR spectra were the same as those without adjusting the pH, confirming that there is no chemical exchange between the protonated form of boric acid and other species.

Method validation

All validation experiments were carried out on a sample of a commercial biocide containing 0.01% of boric acid, previously

analyzed using the ICP-MS technique (used as a reference analytical method).³⁵

The proposed method was validated in order to ensure its reliability during application in routine sample analysis. Thus, several parameters such as selectivity, linear range, trueness (recovery), intra-day precision (repeatability), inter-day precision (intermediate precision), limit of detection (LOD), limit of quantification (LOQ) and uncertainty were evaluated.

Selectivity was estimated analyzing blank samples (D_2O) and commercial biocide samples with concentrations of boric acid below 0.01% (w/w). No signal was observed at the same chemical shift of boron ($\delta_{\text{B}} = 19.4$ ppm), suggesting the absence of matrix interferences that may have given a false positive signal.

The determination of the linear range of a quantitative method is especially crucial for methods with a non-linear detector response. This is not the case for NMR, where the signal is directly proportional to the number of nuclei per volume unit, in the case that the rest of the parameters are constant according to the Curie law (eqn (1)),⁴⁴ where M_0 is the macromagnetization of a system of N spins with spin I and magnetogyric ratio γ , observed in a static magnetic field B_0 at a temperature T .

$$M_0 = \frac{\gamma^2 \hbar I(I+1) B_0}{3K_B T} N \quad (1)$$

In our case, the linear range of the method was confirmed by measuring samples of boric acid (reference standard) at a concentration series of 0.05, 0.10, 0.50, 1.00, 1.50, 3.00 and 4.10% (w/w) in D_2O . The absolute integral of these peaks plotted against the boric acid concentration showed an excellent R^2 linear correlation of 0.9996 (Fig. 2b). Importantly, the excellent correlation is maintained when samples are prepared at the maximum solubility level of 4.1% (w/w) (receiver saturation effects were not observed), proving that the NMR response is linear in the whole range of possible concentrations.

The matrix effect is commonly understood as the difference between the response of the standards in the solvent and those prepared in the matrix similar to the sample to be analyzed. This effect can provoke a signal suppression or enhancement of the analytes because the other components present in the matrix and therefore, the influence of the matrix effect on the response must be studied. The evaluation of the matrix effect involved the comparison of the calibration curve prepared in the solvent, D_2O , with that one prepared by the standard addition calibration method (Fig. 2b), in which increased amounts of the standard analyte are added to sample solutions.⁴⁵ The matrix effect could be then estimated from the slope ratio (SR) of both curves, considering the slope sample/slope D_2O ratio (Fig. 3). The result of 0.97 indicated that there is no matrix effect when measuring boron-11, which reinforced the use of this methodology for quantitative purposes and makes it amenable to determine boric acid in other

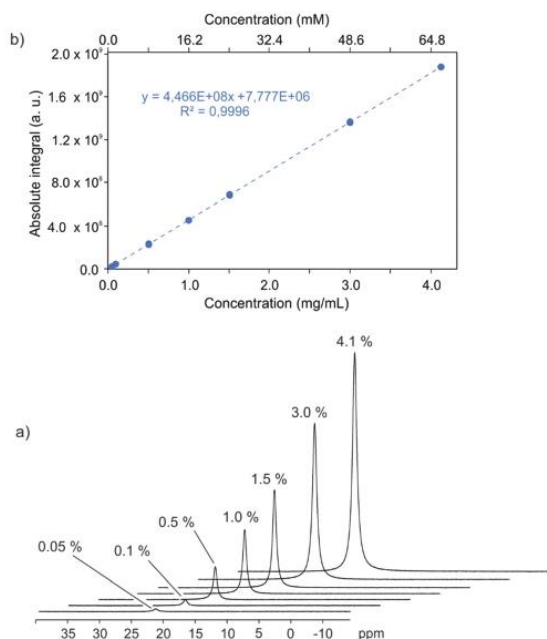


Fig. 2 (a) ^{11}B NMR spectra (96.294 MHz) of boric acid solutions in D_2O at 294 K. Experimental time of each spectrum of 50 seconds; (b) calibration curve stack plot of boric acid solutions in pure D_2O . Average values coming from three replicates are shown. Error bars are hidden behind the spots.

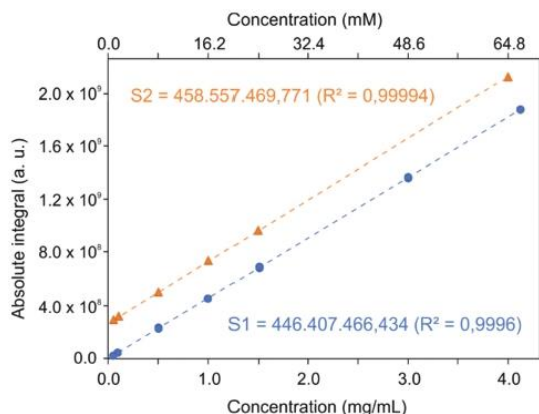


Fig. 3 Calibration curves recorded with solutions of boric acid standard (blue circles, S1 slope) and with extract samples spiked with increasing amounts of boric acid (orange triangles, S2 slope).

matrices such as antiseptics, lubricants, and antifungal or cleaning agents.

Trueness was evaluated in terms of recovery by spiking several aliquots of a commercial sample. The original commercial sample contained 0.01% of boron and for the study, it was spiked with the corresponding addition of a boron analytical standard solution at concentrations of 0.5, 1.5 and 4.0%. Five replicates were tested at each concentration. Average recovery rates obtained at 0.5, 1.5 and 4.0% were 106, 94 and 101%, respectively (see Table 1).

Precision was evaluated by performing intra-day precision (repeatability) and inter-day precision studies, being expressed as relative standard deviation in % (RSD). Intra-day precision was studied by analyzing five spiked samples at three concentration levels (0.5, 1.5 and 4.0%) assayed for the trueness studies and analyzed during the same day. As shown in Table 1, RSD values of 0.9, 1.2 and 0.7% were obtained at 0.5, 1.5 and 4.0%, respectively. The inter-day precision was studied at the same concentration levels but measuring the samples in five consecutive days, obtaining values of 2.0, 1.5 and 1.0%, respectively (Table 1). These precision values are in accordance with the requirements of international guidelines to the studied concentration levels.^{46,47}

LOD and LOQ were determined using the SNR method.^{46,47} Thus, the LOD was set as the lowest concentration of boric

acid that generates an SNR of 3 : 1 and the LOQ as the lowest concentration of boric acid that produces an SNR of 10 : 1, achieving values of 0.02% and 0.04%, respectively (Table 1).

It is important to mention that the SNR is a secondary calculated value and the quantification accuracy of the proposed NMR method must be evaluated by the absolute integrated intensity of the signal, setting the LOQ as the lowest concentration of boric acid that generates suitable precision and accuracy values.⁴⁶ Following the latter statement, the obtained LOQ was 0.04%, which reproduces the one obtained by the SNR method.

Finally, the uncertainty was estimated applying a bottom-up approach.⁴⁸ For this purpose, data obtained during validation at 0.5, 1.5 and 4.0% were used. Moreover, expanded uncertainty (U) was also evaluated and the results are shown in Table 1. For the lower concentration (0.5%), U was 4.4%, whereas for the two higher levels (1.5% and 4.0%), the uncertainty values were 3.9 and 3.7%, respectively.

In order to prove that the developed method can be applied to a wide range of boron compounds, we reproduced the calibration curve using borax (sodium tetraborate) instead of boric acid. All these samples were adjusted at pH 13 in order to ensure that only the monomeric tetraborate species exists in solution. The result of the fit by the linear least squares procedure is shown in Fig. S3.† Good linearity was obtained throughout the tested concentration range, from 0.8 mM (0.05%) to 64.7 mM (4.0%), with a correlation coefficient higher than 0.999.

Sample analysis

Based on the above-mentioned findings, we applied the validated methodology to the quantitative analysis of boric acid in a set of 5 commercial biocides presented as pellets. The spectra of the investigated biocides are shown in Fig. 4. As observed, samples 1, 3 and 5 required two extractions, whereas for samples 2 and 4 the extraction protocol needed to be applied three consecutive times for the complete recovery of the target analyte.

The quantification results of boric acid obtained using the proposed ¹¹B qNMR method are summarized in Table 2. The

Table 1 Validation parameters of the proposed ¹¹B qNMR method

	0.5% (w/w)	1.5% (w/w)	4.0% (w/w)
Recovery (%)	106.2	93.6	101
Intra-day precision ($n = 5$) (%)	0.9	1.2	0.7
Inter-day precision ($n = 5$) (%)	2.0	1.5	1.0
Uncertainty (%)	4.4	3.9	3.7
LOD (%)	0.02		
LOQ (%)	0.04		

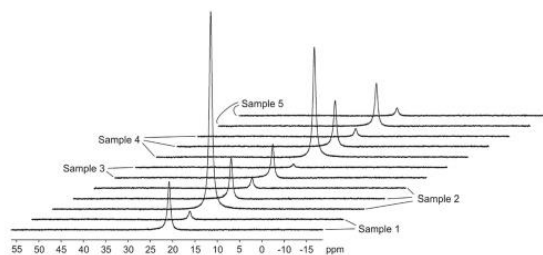


Fig. 4 ¹¹B qNMR (96.294 MHz) spectra of commercial biocide samples (from 1 to 5), prepared as indicated in the experimental section. For samples 1, 3 and 5 two extractions were needed, whereas for samples 2 and 4 three of them were required.

Table 2 Contents of boric acid (qNMR) or elemental boron (ICP-MS) in commercial biocide samples. All concentrations are expressed as % w/w

Samples	Supplier conc.	¹¹ B qNMR	ICP-MS	Diff. (%)
1	2.0	2.1	2.1	0.0
2	10.0	9.9	9.7	2.1
3	1.5	1.6	1.7	5.9
4	9.0–10.0	8.9	9.3	4.3
5	1.0–4.0	2.1	2.0	5.0

concentrations found varied significantly across the different types of investigated products from <LOQ to 10%. In order to check the reliability of the proposed method, the total concentration of boron ions in real samples of commercial biocides was also determined using ICP-MS (Table 2). The differences obtained when compared to our method were always below 6%.

To discard the fact that the ¹¹B-qNMR method is affected by systematic errors, the regression coefficients of both methods were simultaneously evaluated using a joint confidence interval. For ordinary least squares, this joint evaluation can be performed by calculating the *F*-value using eqn (2):

$$F = \frac{n(\gamma - a)^2 + 2(\gamma - a)(\beta - b) \sum x_i + (\beta - b)^2 \sum x_i^2}{2s_E^2} \quad (2)$$

where *n* is the number of data pairs and *s_E²* is the residual variance of the ordinary least squares regression line with slope *b* and intercept *a*. For the joint test, the slope *β* is equal to 1 and the intercept *γ* equals 0. The calculated *F*-value was then compared with the tabulated value for a one-tailed *F* curve with 2 and *n* – 2 degrees of freedom at a significance level of 0.05. The calculated *F*-value was 0.12, which was significantly lower than the tabulated value of 9.55. This result indicates that the ¹¹B qNMR method is not influenced by systematic errors and therefore can be used with confidence for boric acid routine determinations.

Conclusions

A quantitative boron-11 NMR method has been developed for the first time, providing suitable validation parameters such as recovery, precision and uncertainty. Although the LOQ is higher than that for other analytical methods, *i.e.* ICP-MS, this technique is sensitive enough for the reliable determination of boric acid in commercial biocides. This method allows the direct determination of the compound after a simple sample preparation and without requiring large dilution procedures. This quadrupole-based qNMR method offers additional advantages with respect to other chromatographic, spectroscopic and titrimetric methods, such as fast determination (as low as 50 seconds) and the absence of a matrix effect. It is then proposed as a powerful alternative to determine boric acid in complex matrices.

We expect that the methodology described herein will pave the way for the development of new analytical alternatives

based on the acquisition of quadrupolar nuclei. We are currently working on expanding our method for the determination of other boron-containing compounds and in the implementation of other quadrupolar nuclei in qNMR strategies.

Conflicts of interest

The authors declare that they have no conflicting financial interest.

Acknowledgements

This research has been funded by the National R+D+i Plan Project CTQ2017-84334-R of the Spanish Ministry of Science, Innovation and Universities and by the Junta de Andalucía project number P12-FQM-2668. The authors acknowledge the ICP-MS service of the University of Almería.

References

- G. Maniara, K. Rajamoorthi, S. Rajan and G. W. Stockton, Method performance and validation for quantitative analysis by ¹H and ³¹P NMR spectroscopy. Applications to analytical standards and agricultural chemicals, *Anal. Chem.*, 1998, **70**, 4921–4928, DOI: 10.1021/ac980573i.
- J. Cai, J. Zhang, Y. Tian, L. Zhang, E. Hatzakis, K. W. Krausz, P. B. Smith, F. J. Gonzalez and A. D. Patterson, Orthogonal Comparison of GC–MS and ¹H NMR Spectroscopy for Short Chain Fatty Acid Quantitation, *Anal. Chem.*, 2017, **89**, 7900–7906, DOI: 10.1021/acs.analchem.7b00848.
- P. Pregosin, Transition metal nuclear magnetic resonance, *Int. Nucl. Inf. Syst.*, 1991, **13**, 362. https://inis.iaea.org/search/search.aspx?orig_q=RN:24028729.
- P. S. Pregosin, *NMR in Organometallic Chemistry*, Wiley-VCH, Weinheim, Germany, 2012, <https://search.lib.virginia.edu/catalog/u5764567>.
- D. Gudat, *Applications of Heteronuclear X/Y-Correlation Spectroscopy in Organometallic and Organoelement Chemistry: Recent Developments*, 2003, pp. 59–103, DOI: 10.1016/S0066-4103(03)51002-X.
- R. Rigger, A. Rück, C. Hellriegel, R. Sauermoser, F. Morf, K. Breitruck and M. Obkircher, Certified Reference Material for Use in ¹H, ³¹P, and ¹⁹F Quantitative NMR, Ensuring Traceability to the International System of Units, *J. AOAC Int.*, 2017, **100**, 1365–1375, DOI: 10.5740/jaoacint.17-0093.
- D. P. Hollis, Quantitative Analysis of Aspirin, Phenacetin, and Caffeine Mixtures by Nuclear Magnetic Resonance Spectrometry, *Anal. Chem.*, 1963, **35**, 1682–1684, DOI: 10.1021/ac60204a043.
- T. Rundlof, M. Mathiasson, S. Bekiroglu, B. Hakkarainen, T. Bowden and T. Arvidsson, Survey and qualification of



- internal standards for quantification by ^1H NMR spectroscopy, *J. Pharm. Biomed. Anal.*, 2010, **52**, 645–651, DOI: 10.1016/j.jpba.2010.02.007.
- 9 G. F. Pauli, B. U. Jaki and D. C. Lankin, Quantitative ^1H NMR: Development and potential of a method for natural products analysis, *J. Nat. Prod.*, 2005, **68**, 133–149, DOI: 10.1021/np0497301.
 - 10 S. K. Bharti and R. Roy, Quantitative ^1H NMR spectroscopy, *TrAC, Trends Anal. Chem.*, 2012, **35**, 5–26, DOI: 10.1016/j.trac.2012.02.007.
 - 11 G. F. Pauli, T. Gödecke, B. U. Jaki and D. C. Lankin, Quantitative ^1H NMR. Development and potential of an analytical method: An update, *J. Nat. Prod.*, 2012, **75**, 834–851, DOI: 10.1021/np200993k.
 - 12 T. Gödecke, J. G. Napolitano, M. F. Rodríguez-Brasco, S. N. Chen, B. U. Jaki, D. C. Lankin and G. F. Pauli, Validation of a generic quantitative ^1H NMR method for natural products analysis, *Phytochem. Anal.*, 2013, **24**, 581–597, DOI: 10.1002/pca.2436.
 - 13 C. Simmler, J. G. Napolitano, J. B. McAlpine, S. N. Chen and G. F. Pauli, Universal quantitative NMR analysis of complex natural samples, *Curr. Opin. Biotechnol.*, 2014, **25**, 51–59, DOI: 10.1016/j.copbio.2013.08.004.
 - 14 G. F. Pauli, S. N. Chen, C. Simmler, D. C. Lankin, T. Godecke, B. U. Jaki, J. B. Friesen, J. B. McAlpine and J. G. Napolitano, Importance of purity evaluation and the potential of quantitative ^1H NMR as a purity assay, *J. Med. Chem.*, 2014, **57**, 9220–9231, DOI: 10.1021/jm500734a.
 - 15 H. Maki, G. Sakata and M. Mizuhata, Quantitative NMR of quadrupolar nucleus as a novel analytical method: hydrolysis behaviour analysis of aluminum ion, *Analyst*, 2017, **142**, 1790–1799, DOI: 10.1039/c7an00067g.
 - 16 L. S. Simeral, Determination of urea, nitrate, and ammonium in aqueous solution using nitrogen-14 nuclear magnetic resonance, *Appl. Spectrosc.*, 1997, **51**, 1585–1587, DOI: 10.1366/0003702971939145.
 - 17 H. Maki, G. Sakata and M. Mizuhata, Quantitative NMR of quadrupolar nucleus as a novel analytical method: hydrolysis behaviour analysis of aluminum ion, *Analyst*, 2017, **142**, 1790–1799, DOI: 10.1039/c7an00067g.
 - 18 U. Brinkmann-Trettenes, P. C. Stein, B. Klösgen and A. Bauer-Brandl, A method for simultaneous quantification of phospholipid species by routine ^{31}P NMR, *J. Pharm. Biomed. Anal.*, 2012, **70**, 708–712, DOI: 10.1016/j.jpba.2012.07.012.
 - 19 P. Giraudeau, I. Tea, G. S. Remaud and S. Akoka, Reference and normalization methods: Essential tools for the inter-comparison of NMR spectra, *J. Pharm. Biomed. Anal.*, 2014, **93**, 3–16, DOI: 10.1016/j.jpba.2013.07.020.
 - 20 S. Shen, X. Yang and Y. Shi, Application of quantitative NMR for purity determination of standard ACE inhibitors, *J. Pharm. Biomed. Anal.*, 2015, **114**, 190–199, DOI: 10.1016/j.jpba.2015.05.021.
 - 21 Y. Lian, H. Jiang, J. Feng, X. Wang, X. Hou and P. Deng, Direct and simultaneous quantification of ATP, ADP and AMP by ^1H and ^{31}P Nuclear Magnetic Resonance spectroscopy, *Talanta*, 2016, **150**, 485–492, DOI: 10.1016/j.talanta.2015.12.051.
 - 22 N. M. Balsgart, M. Mulbjerg, Z. Guo, K. Bertelsen and T. Vosegaard, High Throughput Identification and Quantification of Phospholipids in Complex Mixtures, *Anal. Chem.*, 2016, **88**, 2170–2176, DOI: 10.1021/acs.analchem.5b03798.
 - 23 B. M. Huang, S. Y. Xiao, T. B. Chen, Y. Xie, P. Luo, L. Liu and H. Zhou, Purity assessment of ginsenoside Rg1 using quantitative ^1H nuclear magnetic resonance, *J. Pharm. Biomed. Anal.*, 2017, **139**, 193–204, DOI: 10.1016/j.jpba.2017.02.055.
 - 24 B. Wrackmeyer, The B-N Bond in Some Aminoboranes and an Iminoborane, Studied by ^{11}B and ^{15}N NMR Spectroscopy and DFT Methods, *Z. Anorg. Allg. Chem.*, 2015, **641**, 2525–2529, DOI: 10.1002/zaac.201500653.
 - 25 B. Wrackmeyer, Organoboranes and tetraorganoborates studied by ^{11}B and ^{13}C NMR spectroscopy and DFT calculations, *Z. Naturforsch., B: Chem. Sci.*, 2015, **70**, 421–424, DOI: 10.1515/znb-2015-0040.
 - 26 J. W. Tomsho and S. J. Benkovic, Elucidation of the Mechanism of the Reaction between Phenylboronic Acid and a Model Diol, Alizarin Red S, *J. Org. Chem.*, 2012, **77**, 2098–2106, DOI: 10.1021/jo202250d.
 - 27 R. A. Oliveira, R. O. Silva, G. A. Moölander and P. H. Menezes, ^1H , ^{13}C , ^{19}F and ^{11}B NMR spectral reference data of some potassium organotrifluoroborates, *Magn. Reson. Chem.*, 2009, **47**, 873–878, DOI: 10.1002/mrc.2467.
 - 28 A. Bagno, W. Kantlehner, R. Kress, G. Saielli and E. Stoyanov, Fries Rearrangement of Aryl Formates: A Mechanistic Study by Means of ^1H , ^2H , and ^{11}B NMR Spectroscopy and DFT Calculations, *J. Org. Chem.*, 2006, **71**, 9331–9340, DOI: 10.1021/jo061475x.
 - 29 M. Bishop, N. Shahid, J. Yang and A. R. Barron, Determination of the mode and efficacy of the cross-linking of guar by borate using MAS ^{11}B NMR of borate cross-linked guar in combination with solution ^{11}B NMR of model systems, *Dalton Trans.*, 2004, 2621–2634, DOI: 10.1039/B406952H.
 - 30 S. Hermanek, ^{11}B NMR Spectra of Boranes, Main-Group Heteroboranes, and Substituted Derivatives. Factors Influencing Chemical Shifts of Skeletal Atoms, *Chem. Rev.*, 1992, **92**, 325–362, DOI: 10.1021/cr00010a007.
 - 31 F. J. Lopez, E. Gimenez and F. Hernandez, Analytical Study on the Determination of Boron in Environmental Water Samples, *Fresenius' J. Anal. Chem.*, 1993, **346**, 984–987, DOI: 10.1007/BF00322763.
 - 32 J. Sciarra and J. Zapotocky, A Study of Methods of Analysis for Boric Acid, *J. Am. Pharm. Assoc.*, 1955, 370–372.
 - 33 M. C. C. Azevedo and A. M. V. Cavaleiro, The acid-base titration of a very weak acid: Boric acid, *J. Chem. Educ.*, 2012, **89**, 767–770, DOI: 10.1021/ed200180j.
 - 34 L. M. Zeng, H. Y. Wang and Y. L. Guo, Fast Quantitative Analysis of Boric Acid by Gas Chromatography-Mass Spectrometry Coupled with a Simple and Selective Derivatization Reaction Using Triethanolamine, *J. Am. Soc.*

- Mass Spectrom.*, 2010, **21**, 482–485, DOI: 10.1016/j.jasms.2009.12.003.
- 35 M. He, Z. Jin, H. Lu and T. Ren, Efficient separation of boron using solid-phase extraction for boron isotope analysis by MC-ICP-MS, *Anal. Methods*, 2015, **7**, 10322–10327, DOI: 10.1039/C5AY01743B.
- 36 P. González, A. Sixto and M. Knochen, Multi-pumping flow system for the determination of boron in eye drops, drinking water and ocean water, *Talanta*, 2017, **166**, 399–404, DOI: 10.1016/j.talanta.2015.12.040.
- 37 Boric acid – substance information –ECHA, (n.d.), <https://echa.europa.eu/es/substance-information/-/substanceinfo/100.030.114> (accessed February 23, 2018).
- 38 R.T.P. HEALTH, Background and environmental exposures to boron in the United States, (n.d.) 9–21, <https://www.atsdr.cdc.gov/toxprofiles/tp26-c2.pdf>.
- 39 World Health Organization, *Boron in Drinking-water*, WHO *Guidel. Drink. Qual.*, 2009, vol. 2, p. 28. http://www.who.int/water_sanitation_health/dwq/boron.pdf.
- 40 E.F.S. Authority, Scientific Opinion on the re-evaluation of boric acid (E 284) and sodium tetraborate (borax) (E 285) as food additives, *EFSA J.*, 2013, **11**, 1–52, DOI: 10.2903/j.efsa.2013.3407.
- 41 Direct analysis of trace elements in gasoline by ICP-MS, (n.d.), <http://tools.thermofisher.com/content/sfs/brochures/D00653~.pdf> (accessed February 23, 2018).
- 42 R. Balz, U. Brändie, E. Kammerer, D. Köhnlein, O. Lutz, A. Nolle, R. Schafitel and E. Veil, 11B and 10B NMR Investigations in Aqueous Solutions, *Z. Naturforsch., A: Phys. Sci.*, 1986, **41**, 737–742, DOI: 10.1515/zna-1986-0508.
- 43 R. C. Weast, *Handbook of Chemistry and Physics*, Cleveland, 1975.
- 44 S. H. Maron and J. B. Lando, *Fundamentals of Physical Chemistry*, Macmillan Publishing, New York, 1974, pp. 209–213.
- 45 J. N. Miller, Basic statistical methods for analytical chemistry. Part 2. Calibration and regression methods. A review, *Analyst*, 1991, **116**, 3–14, DOI: 10.1039/an9911600003.
- 46 U.S. FDA, Guidelines for validation of chemical methods for the FDA foods program, Available from <http://www.fda.gov/downloads/ScienceResearch/FieldScience/UCM298730.pdf>.
- 47 J. Ammon, *et al.*, ‘Guide to NMR Method Development and Validation – Part 1: Identification and Quantification’, EUROLAB Tech. Rep. 1/2014, 2014. <http://www.eurolab.org> (accessed January 18, 2018).
- 48 S. L. R. Ellison and A. Williams, *Eurachem Working Group on Uncertainty in Chemical Measurement, Quantifying uncertainty in analytical measurement*, Eurachem, 2012, <https://www.eurachem.org/index.php/publications/guides/quam> (accessed February 23, 2018).





3.4 Aportación científica 5. Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative ^{31}P nuclear magnetic resonance method

J.R. Belmonte-Sánchez, L.M. Aguilera-Sáez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich. *Microchem. J.* 150 (2019).

DOI: 10.1016/j.microc.2019.104083

CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA DETERMINACIÓN DE TÓXICOS EN AGUAS



Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative ^{31}P nuclear magnetic resonance method



José Raúl Belmonte-Sánchez, Luis Manuel Aguilera-Sáez, Roberto Romero-González, José Luis Martínez Vidal, Francisco Javier Arrebola, Antonia Garrido French*

Department of Chemistry and Physics, Research Centre CIAIMBITAL, Agrifood Campus of International Excellence, ceiA3, University of Almería, Cra. Sacramento, s/n, 04120, Almería, (Spain)

ARTICLE INFO

Keywords:

^{31}P -qNMR
Etidronic acid
Washing water
Analytical validation

ABSTRACT

Etidronic acid (1-hydroxyethylidene-1,1-diphosphonic acid, HEDP) is a common antioxidant additive used as detergent in the processing plants of agri-food industry for fruit and vegetable washing processes before commercialization. In the present work, and for the first time, a quick and reliable alternative to standard established methods for HEDP quantitation study has been performed in vegetable-washing processes applying ^{31}P NMR. The proposed quantitative nuclear magnetic resonance method has been fully validated in water with different analysis times including a novel validation of the limits of detection and quantification depending on the target expected concentration, which results in considerable cost and time savings in those samples containing high concentrations of the compound. The limits of quantification (LOQs) ranged from 0.017% w/v for one minute of analysis (116 scans) to 0.003% w/v for 60 min of analysis (4745 scans). Additional validation parameters have been evaluated such as quantification procedure (comparing internal vs. external calibration), working range (from 0.01% to 0.05% w/v), recovery (95.0%–105.4%) and intra- and inter-day precision (from 0.9 to 4.3%) at 0.01 and 0.05% w/v. The developed method has successfully been applied for the determination of HEDP in 10 samples of agri-food processing industrial waters used for the cleaning of fruits and vegetables prior to their commercialization, providing adequate results and obtaining HEDP concentrations between 0.003 and 0.050% w/v.

1. Introduction

Biphosphonates, such as etidronic acid (HEDP), are extensively use, as other phosphonates, to prevent mineral precipitation and inhibiting corrosion in industrial washing and cooling systems [1]. They also act as scale inhibitors because they are excellent complexation agents of heavy metals [2]. Phosphonates are very stable against biological degradation and they were found to be active at concentrations higher than 100 mg/L (0.01%), causing moderate acute oral and dermal toxicity on humans [3]. Specifically, HEDP has an excellent stability in aquatic ecosystems due to its high persistence against microbial degradation and extraordinary resistance to hydrolysis [1].

Several HEDP determinations has been previously reported applying different techniques as inductively coupled plasma mass spectrometry (ICP-MS) [4], ion chromatography with inverse UV detection [5], and capillary electrophoresis with indirect photometric detection [2], but they are very time-consuming techniques and large amounts of solvents are required for compound extraction.

Previous studies of different compounds containing phosphorus by quantitative phosphorus NMR (^{31}P -qNMR) have been reported for the assessment of impurities in highly concentrated samples but no strict validation methods have been achieved [2,6–8]. Certain advantages make this nucleus suitable for quantitative analysis by NMR: 31-phosphorus isotope is 100% natural abundance in the nature, it is a high sensitive nuclei with spin $I = 1/2$ and presents a wide chemical shift range allowing excellent separation of the signals. To the best of our knowledge, ^{31}P -qNMR has been limited to applications for quantifying analytes such as phosphorus-containing pesticides [9], phosphonates [2], phosphorylated amino acids [10] or phospholipid species [11] in natural matrices.

Over the years, qNMR has gained popularity in the analysis of different kind of matrices, such as complex natural products, dietary supplements, biocide tablets or traditional medicines [12–17]. Besides, a great improvement has been achieved in the electronics and materials of NMR spectrometers. NMR has always been considered a very limited quantitative technique due to its lower sensitivity compared to others

* Corresponding author.

E-mail address: agarrido@ual.es (A. Garrido French).

<https://doi.org/10.1016/j.microc.2019.104083>

Received 23 February 2019; Received in revised form 2 June 2019; Accepted 8 July 2019

Available online 09 July 2019

0026-265X/ © 2019 Elsevier B.V. All rights reserved.

common analytical techniques, but these technological advances allow improving sensitivity performance of NMR in relation to well-established techniques such as gas chromatography (GC) and liquid chromatography (LC) coupled to mass spectrometry (MS), which have been extensively applied for the detection of all kind of molecules in microgram per kilogram (ppb) or nanogram per kilogram (ppt) levels. However, these are very time consuming including complex sample treatments and demand large volumes of costly solvents [18,19].

On the other hand, NMR demands minimal sample preparation and allows direct sample analysis. Also, this technique is non-destructive and provides the detection and structural determination of different compounds in a single experiment, even if the sample is a complex mixture or the sample solution contains a solid phase [20]. As a result, high-accuracy quantitative NMR (qNMR) allows for quantifying unknown compounds in mixtures by comparing its relative integration area to that obtained from a known amount of a standard or using specific calibration curves, in addition to the advantages of NMR spectroscopy [21]. For these reasons, NMR spectroscopy is nowadays a reliable analytical tool and it has become itself as a routine technique [11].

For qNMR calibration, both internal and external calibration methods can be performed, and the first one is the most used for individually qNMR determinations if the sample characteristics lead to optimal conditions (e.g non-overlapping resonances or adequate relaxation delays, D1). In addition, the internal calibration method only requires one single calibration standard for quantifying multiple analytes among the same analysis [22]. Alternatively, an external calibration can be performed as for many other analytical techniques such as spectrometry [23,24]. Under certain circumstances, this external calibration method can lead to a considerable simplification of sample preparation because no addition of internal standard in each sample is necessary and large batches of samples can be analyzed by using the same calibration curve. In addition, the absence of the matrix effect allows the application in different matrices [25]. Potential differences (pH, ionic strength, etc.) must be correctly controlled by buffer solutions to ensure the reproducibility of the calibration curve and its application to the analysis of targeted samples [11].

The validation of quantitative methods using qNMR is usually based on guidelines focused on purity assay of pharmaceuticals and reference materials and on the determination of presence of impurities, degradation products and matrix components [14,26–30]. Some key validation aspects such as calculation of the LOQ of qNMR methods are not properly defined yet, and considering the lack of validation guidelines for qNMR analytical methods, its application represents a challenge.

In the present study, ^{31}P -qNMR has been applied for the determination of low concentrations of HEDP in industrial waters of vegetable-washing processes before their commercialization. Moreover, a fully validation of a rapid, sensitive, and reliable qNMR-based method has been performed, taking advantage in comparison with alternative techniques such as LC-MS. The developed method was also successfully applied to the analysis of real water samples collected from agri-food processing industries.

2. Materials and methods

2.1. Materials and reagents

HEDP ($\text{C}_2\text{H}_8\text{O}_7\text{P}_2$, purity > 95%), potassium chloride (purity > 99%) and potassium hydrogen phthalate (purity > 99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Sodium hydroxide was obtained from Panreac (Barcelona, Spain) and di-sodium hydrogen phosphate anhydrous was purchased from Merck (Darmstadt, Germany). Deuterium oxide (D_2O , $\text{D} \geq 99.9\%$) and Norell™ 5.0 mm O.D. \times 178 mm borosilicate NMR sample tubes were purchased from Eurisotop (St-Aubin Cedex, France). Milli Q water (Millipore, Bedford,

MA, USA) was used for method optimization and calibration.

Materials and reagents applied for liquid chromatography-mass spectrometry (LC-MS) experiments are described in Supplementary Material.

2.2. Apparatus

NMR spectra were obtained on a Bruker Avance 300 MHz (Bruker Company, Switzerland) using a direct BBO^{PLUS} (^1H , BB- ^{19}F , SmartProbe) probehead and a Bruker Avance III HD 600 MHz spectrometer equipped with a QCI $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{31}\text{P}$ proton-optimized quadrupole inverse cryoprobe with ^1H and ^{13}C cryochannels. Chemical shifts (δ) are given in parts per million (ppm). ^{31}P chemical shifts are given relative to Na_2HPO_4 ($\delta_p = 3.1$ ppm).

The LC-MS instrument details are summarized in Supplementary Material.

2.3. Samples and sample preparation

A total of 10 vegetable-washing water samples were supplied from different local agri-food industries (Almería, Spain) for their study by LC-MS and ^{31}P NMR. Before NMR measurements, water samples were treated as follows: 1 mL of the representative sample was adjusted in an Eppendorf tube at pH 11 by adding 500 μL of $\text{Na}_2\text{HPO}_4/\text{NaOH}$ 0.25 M buffer solution. 200 μL of deuterated water (D_2O) were added and vortexed at room temperature for 5 min. Then, 1 mL of the resulting solution was directly transferred to the NMR tube, which was sealed prior to analysis.

2.4. ^{31}P -qNMR method

^{31}P -qNMR spectra were recorded at an operating frequency of 121.495 MHz and a spectral width of 48.077 kHz (395.71 ppm) using standard Bruker method (zsig 1D sequence with inverse gated decoupling). Additional analyses by 600 MHz spectrometer were performed, operating at 242.937 MHz in the same conditions. Shimming was performed on each sample prior to data acquisition using the TopShim automatic shimming method from Bruker BioSpin. All samples were measured at 20.0 ± 0.1 °C and under nonspinning conditions with regard to the high magnetic field to avoid spinning side bands. NMR spectra were collected using the following acquisition parameters: size of fid (TD) = 65,536, acquisition time (AQ) = 0.68 s, relaxation delay (D1) = 0.05 s, variable number of scans (NS) from 116 to 4745, pulse width (P1) = 12 μs and a 90° pulse angle. The Lorentzian line-broadening factor of 10 Hz was applied to the total FID prior to Fourier transformation. The spectra were automatically phased, baseline-corrected and integrated using TOPSPIN software (version 4.0). The spectrometer transmitter was locked to D_2O frequency throughout each analysis. All NMR experiments were performed with a fixed receiver gain (RG) of 203, which was the optimum value estimated through several tests.

2.5. Method validation

The proposed ^{31}P -qNMR method for HEDP determination was validated in terms of selectivity, linear working range from 0.01% to 0.05% w/v, intra-day, inter-day precision, LOQ and trueness (recovery).

Trueness was calculated by estimating the recovery at two different concentration levels (0.01% and 0.05% w/v). Thus, five replicates were analyzed five times in the same day (intraday precision or repeatability) and in five consecutive days (interday precision or reproducibility). The LOQ of the proposed method was estimated as the lowest concentration of HEDP that generates suitable precision and accuracy values (95%–105%).

3. Results and discussion

3.1. Preliminary LC-MS analysis

To our knowledge, no previous HEDP determination in aqueous samples by LC-MS methods have been published, despite nowadays this is the golden technique for the determination of polar compounds as HEDP. Therefore, a series of preliminary tests were carried out for attempting the HEDP determination by LC-MS in order to check the possibility of using this technique as analytical reference method. Firstly, three transitions were selected to monitor the compound, showing the selected ions in Table S-1 (see Supplementary Material).

However, when the aqueous samples (see Section 2.3) were directly injected into the LC system (see supplementary material for further details regarding experimental conditions), no signal of the compound was obtained, neither in the samples nor fortified samples with the target compound. This fact can be explained by high suppression matrix effect because the high polarity of the compounds and its low retention in the stationary phase. Alternative strategies were used to increase the retention of the compound or to reduce its polarity: a) different columns, which have been utilized for the determination of polar compounds [31], were tested; b) fluorenylmethyloxycarbonyl chloride (FMOC-Cl) was used as derivatization reagent, bearing in mind that it has been widely used for similar compounds as glyphosate [32]; c) solid phase extraction has also been included to remove some compounds from the matrix. Fortified samples were analyzed using these alternative procedures and again, no chromatographic signal was obtained for HEDP. Bearing in mind the high polarity of the target compound as well as the complexity of the studied samples, it was concluded that the tested approaches by LC-MS do not provided suitable results for the determination of HEDP in aqueous samples.

3.2. qNMR method optimization

Considering that the molecule contains two phosphorus atoms with nuclear magnetic resonance, ^{31}P NMR was selected instead of chromatographic techniques. The proposed methodology could be an interested approach because the high sensitivity of phosphorous-31 (spin nucleus 1/2, 100% abundance) and high chemical shift dispersion. In addition, this nucleus produces sharp signals, so line shapes are less affected by the inhomogeneity of the magnetic field [2].

Initially, an appropriate deuterated solvent was needed for adjusting the lock frequency. Due to the aqueous nature of the washing samples and consequently HEDP water solubility, deuterated water was

selected as an appropriate choice for the optimisation of the method. One single signal corresponding to HEDP (δ_{p} 20.0 ppm) was obtained in the initial $^{31}\text{P}\{^1\text{H}\}$ -NMR experiments (zgig 1D sequence with inverse gated decoupling) using a standard solution (0.02% w/v), which provided a significant signal. For all qNMR experiments, a 90° pulse flip angle was selected leading to higher SNR than smaller pulse angles with the same number of scans (SNR and the square root number of scans are directly proportional) [33].

The pH of the sample also plays a significant role in qNMR analysis. Differences in the pH of water samples may cause drifts in the chemical shift and widening of the desired signals [14]. To evaluate the optimal pH for the HEDP determination, three different experiments using the 0.02% w/v standard solution were carried out at the pH values of 2.3, 6.5 and 11. Initially, the pH optimization experiments were adjusted by adding 500 μL of sodium hydroxide/potassium chloride 0.25 M buffer and potassium hydrogen phthalate/hydrochloric acid 0.25 M buffer for pH 11 and pH 2.3, respectively. Whilst the small change in chemical shift ($\Delta\delta_{\text{p}}$) between the different pH values (Fig. S-1) suggests that the dissociation of protons on the phosphonic acid groups has a small effect on the shielding of the phosphorus nuclei [34], and the most intense and sharpest signal was obtained for the alkaline tested solution (δ_{p} 19.9 ppm).

With the aim of HEDP quantification, the internal standard method was initially evaluated. For that, Na_2HPO_4 was used as internal quantification standard due to the fact that it is a common compound employed for ^{31}P -qNMR [8]. Besides, it fulfils specific requirements such as high purity, good stability and solubility in the selected solvent, low volatility, chemical inertness with the analyte of interest, and finally it provides a sharp and separated signal that does not interfere with the signal of the HEDP in the NMR spectra. Thus, a $\text{Na}_2\text{HPO}_4/\text{NaOH}$ 0.25 M buffer solution in D_2O water allowed both pH adjusting (pH 11) and internal standard quantification at the same time without signal overlapping. Chemical shifts were relative to the internal standard (IS) signal at δ_{p} 3.1 ppm.

The D1 time, which describes the time needed by the nucleus to go back to ground state after the excitation pulse [35], was optimized by progressive D1 variation experiments of HEDP and IS solutions. This speed is influenced by the chemical surroundings of the detected ^{31}P nucleus. Improper D1 values could result in lower signal intensity than the true value or in long experimental processes and higher costs. Therefore, both analyte and IS phosphorous nucleus relaxation time were studied using a solution of HEDP (0.02% w/v) and 500 μL of the buffer solution. D1 was ranged from 0.05 s to 60 s and the intensities obtained for HEDP and IS are shown in Fig. 1. It can be observed that at

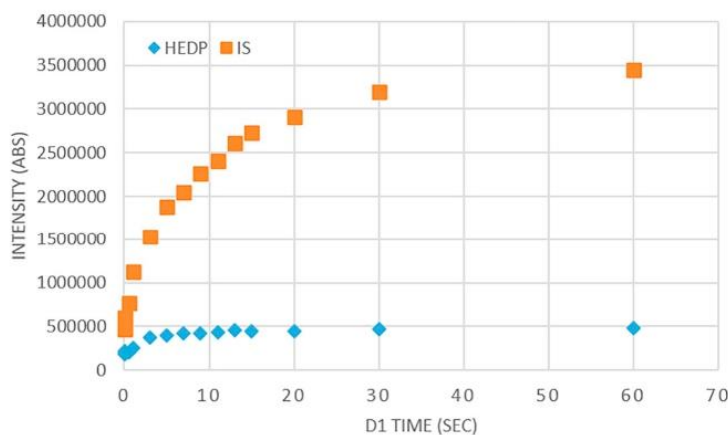


Fig. 1. Determination of the optimal relaxation delay (D1) by monitoring of the intensity of the HEDP (◆) and IS ^{31}P NMR signal (■) vs. progressive D1 time increasing.

least, a minimum D1 = 10 s for HEDP and a minimum of 60 s for IS were necessary for obtaining an optimal relaxation, which is in accordance with previous reported data [8,36].

Taking into account the lack of water-soluble primary phosphorous IS materials with shorter D1 relaxation times (i.e. phosphonacetic acid, D1 > 23 s) and the aim of reaching low LOQ of the target compound in a rapid NMR method, an external calibration method was additionally evaluated for HEDP. In order to determine the optimal relationship between the number of scans, and intensity of the obtained HEDP signal for the proposed NMR method, different D1 times from 0.01 to 10 s were evaluated. Within the same time of analysis of 10 min and the corresponding variation in the number of scans (from 5462 scans for D1 = 0.01 s to 52 scans for D1 = 10 s), the most intense signal was obtained at D1 = 0.05 s. Thus, applying this D1, although no maximum integral area was obtained at this relaxation time, the external calibration provides an adequate working range where the integral areas can be related to concentrations with optimal linearity as long as the same acquisition and processing parameters are used (number of scans, pulse angle, etc.), by using this method, the analysis time is considerably short, allowing the accumulation of greater number of scans. Finally, to evaluate the optimal NS, they were evaluated ranging from 116 (total time of analysis: 1 min) to 4745 scans (60 min) using the lowest concentration of the HEDP working range (0.01% w/v). As expected, an adequate linear correlation between signal area and NS was found because of the high repeatability and accuracy of the radio frequency transmitter and pulse generator (Fig. 2A).

Typical concentrations of HEDP found in real samples (around 0.02% w/v) require a NS value of 232 scans (2 min) to afford quantitative results, which results in a fast method that can provide adequate LOQs. Nevertheless, if lower LOQ is required, NS parameter can be increased significantly reducing drastically the LOQ of the method as can be seen in Fig. 2B where the signal-to-noise ratio (SNR) for the HEDP is proportional to square root of the number of scans.

As well as the optimization of data acquisition, the post-processing parameters are also important for improving the accuracy and precision of the method. Each spectrum was calibrated manually previous to its processing to the phosphate signal at δ_P 3.1 ppm. Before Fourier transformation the ^{31}P spectra were processed with a line broadening of 10 Hz. The automatic baseline corrections were adjusted manually when necessary to ensure the correct integration of the signal. With this aim, each signal was manually integrated five times and the average of the obtained value was used for calculations.

3.3. Method validation

The proposed method was validated considering a compilation of different established guidelines that include the following sections and assuring an optimal analytical validation parameters as selectivity, working range, precision, LOQ, stability, trueness and uncertainty [37–41].

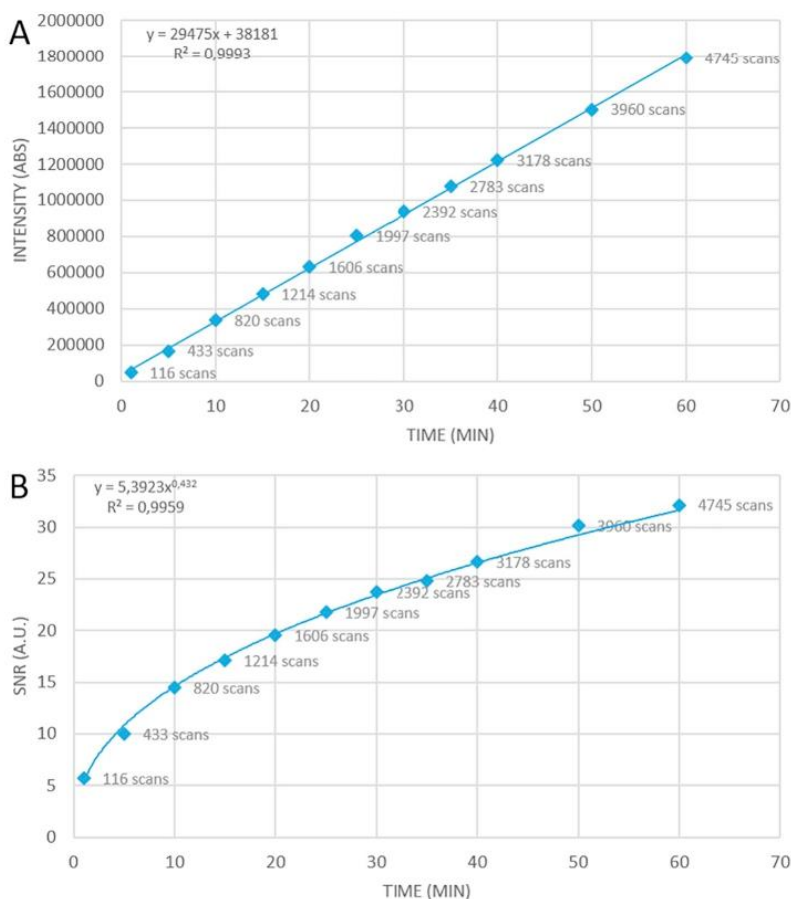


Fig. 2. A) Influence of number of scans on the absolute signal integral. B) Influence of number of scans on the signal-to-noise ratio.

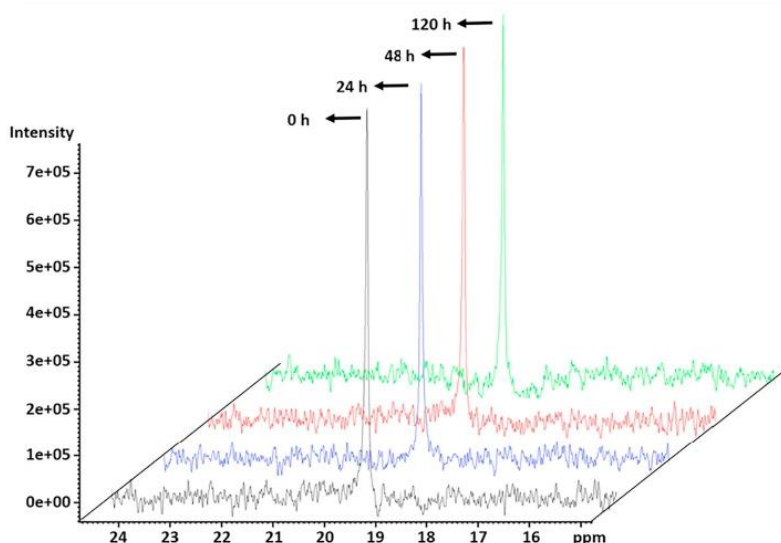


Fig. 3. Stability determination plot for a 0.01% HEDP solution at 0 (black), 24 (blue), 48 (red) and 120 h (green) at ambient temperature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3.1. Selectivity

Selectivity was studied to estimate the possible interference or impurities from sample solutions. With this aim, blank samples (H_2O) and industrial water samples were analyzed. Potential signals were searched at the chemical shift dispersion of phosphorous nucleus and it was not observed any interference that could have been confused with those obtained for the studied compound or calibration standards. Therefore, a satisfying selectivity was obtained for the proposed qNMR method when it was applied to the analysis of fruits and vegetable washing waters.

3.3.2. Stability

The stability test of HEDP was conducted at 0, 24, 48 and 120 h by measuring of 3 replicates of a standard solution (0.01% w/v) at room temperature ($25 \pm 2^\circ C$) and compared with initial values (Fig. 3). The stability results were satisfactory obtaining a RSD value of 1.43% and ensuring the stability of the target compound in aqueous solutions during periods of at least 5 days.

3.3.3. Working range

The working range of the calibration curve for HEDP was determined by triplicate using three different concentrations that ranged from 0.01% to 0.05% w/v (Table 1). The standards were diluted with

Table 1
Validation parameters of the developed method.

Tested parameter	HEDP concentration		
	0.01% (w/v)	0.02% (w/v)	0.05% (w/v)
Intra-day precision (n = 3) (%)	1.58	0.99	0.88
Inter-day precision (n = 3) (%)	4.30	2.22	1.19
Recovery (%)	94.3–105.4	–	95.0–98.4
Uncertainty (%)	8.9	–	3.7
LOQ	0.01% w/v (1 min, 116 scans) - 0.003% w/v (60 min, 4745 scans)		
Linearity	$R^2 = 0.997$		
Stability	1.43% (120h)		

Milli Q water. The linearity of the calibration curve was estimated by the use of the determination coefficient (R^2), showing a good linearity ($R^2 = 0.997$) in the studied range (Fig. 4). Simple water dilutions allow to analyse samples containing higher concentrations than the upper level of the validated working range.

3.3.4. Precision

Precision was estimated by performing intra and inter-day experiments. The precision results were expressed in terms of relative standard deviation (RSD) at two different concentration levels (0.01% and 0.05% w/v). The obtained values for precision of the present method are in accordance with the requirements of international guidelines to the studied concentration levels (Table 1) [37,42].

3.3.5. LOQ

The LOQ, understood as the lowest concentration of the compound that produces an SNR of 10:1, was usually determined. Nevertheless, it is important to indicate that the SNR is a secondary calculated value and the quantification accuracy of the proposed NMR method must be evaluated by the absolute integrated intensity of the signal, setting the LOQ as the lowest concentration of HEDP that generates suitable precision and accuracy values [43]. Following the latter statement, the calculated LOQ for a sample at 0.01% w/v concentration ranged from 0.01% w/v, for one minute of analysis (116 scans) to 0.003% w/v for 60 min of analysis (4745 scans).

An additional study for calculation of the LOQ of the method using an Avance III HD 600 MHz spectrometer was performed to demonstrate the scope of the method when a higher sensitivity equipment is used. With this aim, a 60 min (4745 scans) analysis of a sample at 0.01% w/v concentration was performed, obtaining in this case an improved LOQ of 0.002% w/v.

3.3.6. Trueness and uncertainty

Trueness, defined as the agreement between the measured value and the true value [25], was tested in blank industrial water by the standard addition technique at the higher and lower limit of the working range (0.01% to 0.05% w/v). The results were expressed in terms of recovery (%). Acceptable recovery values, ranging from 95.0% to 105.4%, were obtained for both concentrations studied (Table 1).

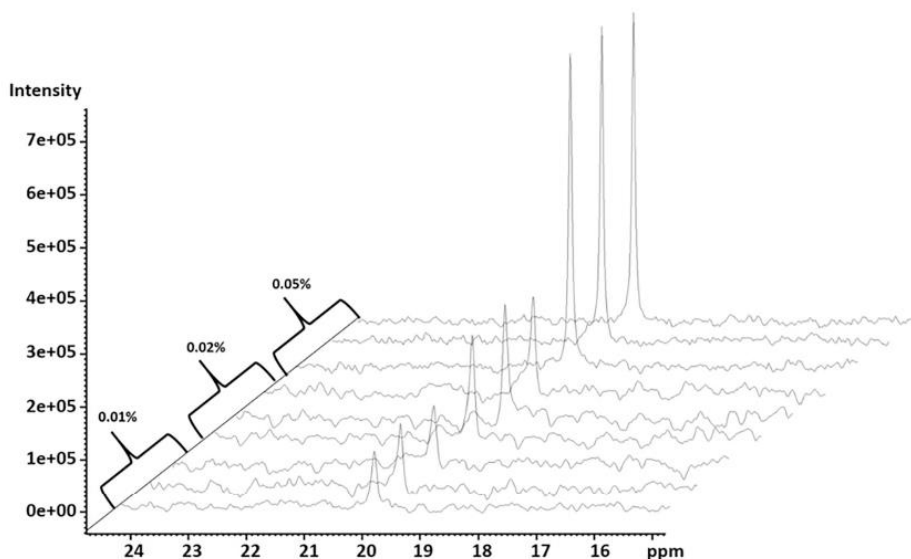


Fig. 4. Calibration curve stack plot of HEDP solutions in pure D_2O obtained for three replicates of 0.01, 0.02 and 0.05% (w/v), respectively.

Finally, the uncertainty (U) was estimated applying a bottom-up approach [39]. This calculation is commonly estimated during the validation of reliable analytical methods, but it has hardly been documented in previously qNMR published articles. For this purpose, data obtained during validation at 0.01 and 0.05% w/v were used. For the lowest tested concentration (0.01% w/v), U was 8.9% whereas for the highest level (0.05% w/v), the calculated uncertainty value was 3.7%, indicating the suitability of the proposed method for quantification measurements.

3.3.7. Real samples

The developed method was applied for the determination of HEDP in 10 samples of different agro-food processing industrial waters used for the cleaning of fruits and vegetables prior to commercialization. HEDP was found in one of the studied samples at the concentration of 0.027% w/v and also it was detected in another sample at a concentration below the LOQ of the method ($< 0.01\%$ w/v) although appreciable signal was observed, as it can be observed in Fig. S-2, where the spectra of the samples are shown. To achieve a quantifiable signal in the optimized work range of this sample, a longer number of scans have been necessary. With this objective, an analysis with 2783 scans (40 min of acquisition) allowed us to calculate the exact concentration of HEDP in the sample, determined at 0.008% w/v.

4. Conclusions

This paper describes for the first time the application of ^{31}P -qNMR to determine low concentrations of HEDP in industrial waters used for vegetable and fruits washing processes before commercialization. The proposed analytical method represents clear advantage in relation to alternative techniques in terms of simplicity, speed and sensitivity and overcome the limitations observed by LC-MS. The method was properly validated according to different established validation guidelines and verified the usefulness of external calibration applied to the analysis of 31-P containing compounds such as HEDP, improving the quantitative results regarding the use of the internal standard method using Na_2HPO_4 as internal quantification standard. The validation showed suitable selectivity, linearity in the working range, precision, LOQ, stability, trueness and uncertainty.

The developed method is sensitive enough for the reliable determination of HEDP at different concentration ranges, achieving several LOQs depending on the number of scans and sensitivity required, ranging from 0.01% w/v, for one minute of analysis (116 scans) to 0.003% w/v for 60 min of analysis (4745 scans).

A total of 10 real samples were analyzed using the developed method, obtaining positive results for two of them (0.027% and 0.008% w/v). For the sample with lower concentration, a larger number of scans were applied to achieve a quantifiable signal within the working range. Furthermore, simple water dilutions allow to analyse samples containing higher concentrations than the validated working range.

The method has demonstrated to be appropriate for HEDP routine determinations in different NMR spectrometers, obtaining adequate results through a fast, reliable method and without the need for complex sample treatment with the consequent reduction of time and costs. We believe that methodologies such as the one proposed here for the validation of qNMR methods will mark the way for the development of new analytical alternatives.

Declaration of Competing Interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.microc.2019.104083>.

References

- [1] E. Rott, H. Steinmetz, J.W. Metzger, Organophosphonates: a review on environmental relevance, biodegradability and removal in wastewater treatment plants, *Sci. Total Environ.* 615 (2018) 1176–1191, <https://doi.org/10.1016/j.scitotenv.2017.09.223>.
- [2] M. Oromí-Farrús, J.M. Minguell, N. Oromí, R. Canela-Garayoa, A reliable method for quantification of phosphonates and their impurities by ^{31}P NMR, *Anal. Lett.* 46 (2013) 1910–1921, <https://doi.org/10.1080/00032719.2013.780239>.
- [3] Safety Assessment of Etidronic Acid and its Simple Salts as Used in Cosmetics, <https://www.cir-safety.org/sites/default/files/Etidronicacid.pdf>, (2016), Accessed date: 7 November 2018.
- [4] C.K. Schmidt, B. Raue, H.J. Brauch, F. Sacher, Trace-level analysis of phosphonates in environmental waters by ion chromatography and inductively coupled plasma

- mass spectrometry, *Int. J. Environ. Anal. Chem.* 94 (2014) 385–398, <https://doi.org/10.1080/03067319.2013.831410>.
- [5] J. Zirojevic, Z. Jovic, A. Djurdjevic, A. Ciric, P. Djurdjevic, Chemometric-assisted determination of some bisphosphonates and their related substances in pharmaceutical forms by ion chromatography with inverse UV detection, *Acta Chromatogr.* 27 (2015) 215–237, <https://doi.org/10.1556/ACHrom.27.2015.2.2>.
- [6] K. Moedritzer, L. Maier, L.C.D. Groeneweghe, Phosphorus-31 nuclear magnetic resonance spectra of phosphorus compounds, *J. Chem. Eng. Data* 7 (1962) 307–310, <https://doi.org/10.1021/jc60013a043>.
- [7] D.R. Gard, J.C. Burquin, J.K. Gard, Quantitative analysis of short-chain phosphates by phosphorus-31 nuclear magnetic resonance and interlaboratory comparison with infrared and chromatographic methods, *Anal. Chem.* 64 (1992) 557–561, <https://doi.org/10.1021/ac00029a020>.
- [8] M. Weber, C. Hellriegel, A. Rueck, J. Wuethrich, P. Jenks, M. Obkircher, Method development in quantitative NMR towards metrologically traceable organic certified reference materials used as ³¹P NMR standards, *Anal. Bioanal. Chem.* 407 (2015) 3115–3123, <https://doi.org/10.1007/s00216-014-8306-6>.
- [9] R. Rigger, A. Rück, C. Hellriegel, R. Sauermoser, F. Morf, K. Breitruck, M. Obkircher, Certified reference material for use in ¹H, ³¹P, and ¹⁹F quantitative NMR, ensuring traceability to the international system of units, *J. AOAC Int.* 100 (2017) 1365–1375, <https://doi.org/10.5740/jaoacint.17-0093>.
- [10] A. Stücker, J. Podschun, B. Saake, R. Lehnen, A novel quantitative ³¹P NMR spectroscopic analysis of hydroxyl groups in lignosulfonic acids, *Anal. Methods* 10 (2018) 3481–3488, <https://doi.org/10.1039/C8AY01272E>.
- [11] U. Brinkmann-Trettenes, P.C. Stein, B. Klösgen, A. Bauer-Brandl, A method for simultaneous quantification of phospholipid species by routine ³¹P NMR, *J. Pharm. Biomed. Anal.* 70 (2012) 708–712, <https://doi.org/10.1016/j.jpba.2012.07.012>.
- [12] T. Rundlof, M. Mathiasson, S. Bekiroglu, B. Hakkarainen, T. Bowden, T. Arvidsson, Survey and qualification of internal standards for quantification by ¹H NMR spectroscopy, *J. Pharm. Biomed. Anal.* 52 (2010) 645–651, <https://doi.org/10.1016/j.jpba.2010.02.007>.
- [13] H.S. Jang, B. Jeong, S.Y. Choi, G.H. Jang, K.C. Park, Y.S. Kwon, H. Yang, Conditrol F, the discriminant marker between *C. wilfordii* and *C. auriculatum* by ¹H NMR spectroscopy, *Microchem. J.* 135 (2017) 153–157, <https://doi.org/10.1016/j.microc.2017.09.001>.
- [14] S.K. Bharti, R. Roy, Quantitative ¹H NMR spectroscopy, *TrAC Trends Anal. Chem.* 35 (2012) 5–26, <https://doi.org/10.1016/j.trac.2012.02.007>.
- [15] C. Simmler, J.G. Napolitano, J.B. McAlpine, S.N. Chen, G.F. Pauli, Universal quantitative NMR analysis of complex natural samples, *Curr. Opin. Biotechnol.* 25 (2014) 51–59, <https://doi.org/10.1016/j.copbio.2013.08.004>.
- [16] A. Owczarek, A. Klys, M.A. Olszewska, A validated ¹H qNMR method for direct and simultaneous quantification of esculin, fraxin and (–)-epicatechin in Hippocastani cortex, *Talanta* 192 (2019) 263–269, <https://doi.org/10.1016/j.talanta.2018.09.036>.
- [17] W. Wollinger, J.L.N. Fernandes, L.H.K. Queiroz Júnior, B.C. Garrido, F.R. de Aquino Neto, Improving quantitative ¹³C NMR performance by an adiabatic scheme, *Microchem. J.* 140 (2018) 167–175, <https://doi.org/10.1016/j.microc.2018.04.021>.
- [18] H.M. McNair, J.M. Miller, *Basic Gas Chromatography*, Second Edn, John Wiley & Sons, Inc, New Jersey, 2009.
- [19] J. Cai, J. Zhang, Y. Tian, L. Zhang, E. Hatzakis, K.W. Krausz, P.B. Smith, F.J. Gonzalez, A.D. Patterson, Orthogonal comparison of GC–MS and ¹H NMR spectroscopy for short chain fatty acid quantitation, *Anal. Chem.* 89 (2017) 7900–7906, <https://doi.org/10.1021/acs.analchem.7b00848>.
- [20] H. Maki, G. Sakata, M. Mizuhata, Quantitative NMR of quadrupolar nucleus as a novel analytical method: hydrolysis behaviour analysis of aluminum ion, *Analyst* 142 (2017) 1790–1799, <https://doi.org/10.1039/c7an00067g>.
- [21] L.M. Aguilera-Sáez, J.R. Belmonte-Sánchez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich, I. Fernández, Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides, *Analyst* 143 (2018) 4707–4714, <https://doi.org/10.1039/C8AN00505B>.
- [22] G.F. Pauli, T. Gödecke, B.U. Jaki, D.C. Lankin, Quantitative ¹H NMR. Development and potential of an analytical method: an update, *J. Nat. Prod.* 75 (2012) 834–851, <https://doi.org/10.1021/np200993k>.
- [23] P. Giraudeau, Quantitative 2D liquid-state NMR, *Magn. Reson. Chem.* 52 (2014) 259–272, <https://doi.org/10.1002/mrc.4068>.
- [24] C.I. Cerceau, L.C.A. Barbosa, E.S. Alvarenga, A.G. Ferreira, S.S. Thomasi, A validated ¹H NMR method for quantitative analysis of α-bisabolol in essential oils of *Eremanthus erythropappus*, *Talanta* 161 (2016) 71–79, <https://doi.org/10.1016/j.talanta.2016.08.032>.
- [25] P. Giraudeau, Challenges and perspectives in quantitative NMR, *Magn. Reson. Chem.* 55 (2017) 61–69, <https://doi.org/10.1002/mrc.4475>.
- [26] G.F. Pauli, S.N. Chen, C. Simmler, D.C. Lankin, T. Gödecke, B.U. Jaki, J.B. Friesen, J.B. McAlpine, J.G. Napolitano, Importance of purity evaluation and the potential of quantitative ¹H NMR as a purity assay, *J. Med. Chem.* 57 (2014) 9220–9231, <https://doi.org/10.1021/jm500734a>.
- [27] X. Li, K. Hu, Quantitative NMR Studies of Multiple Compound Mixtures, (2017), <https://doi.org/10.1016/bs.ammr.2016.08.001>.
- [28] D.W.H. Merx, Y. Westphal, E.J.J. van Velzen, K.V. Thakoer, N. de Roo, J.P.M. van Duynhoven, Quantification of food polysaccharide mixtures by ¹H NMR, *Carbohydr. Polym.* 179 (2018) 379–385, <https://doi.org/10.1016/j.carbpol.2017.09.074>.
- [29] G.G. Shimamoto, L.F. Bianchessi, M. Tubino, Alternative method to quantify biodiesel and vegetable oil in diesel-biodiesel blends through ¹H NMR spectroscopy, *Talanta* 168 (2017) 121–125, <https://doi.org/10.1016/j.talanta.2017.03.027>.
- [30] E. Kang, H.R. Park, J. Yoon, H.Y. Yu, S.K. Chang, B. Kim, K. Choi, S. Ahn, A simple method to determine the water content in organic solvents using the ¹H NMR chemical shifts differences between water and solvent, *Microchem. J.* 138 (2018) 395–400, <https://doi.org/10.1016/j.microc.2018.01.034>.
- [31] R. Domingos Alves, R. Romero-González, R. López-Ruiz, M.L. Jiménez-Medina, A. Garrido Frenich, Fast determination of four polar contaminants in soy nutraceutical products by liquid chromatography coupled to tandem mass spectrometry, *Anal. Bioanal. Chem.* 408 (2016) 8089–8098, <https://doi.org/10.1007/s00216-016-9912-2>.
- [32] A.M. Botero-Coy, M. Ibáñez, J.V. Sancho, F. Hernández, Improvements in the analytical methodology for the residue determination of the herbicide glyphosate in soils by liquid chromatography coupled to mass spectrometry, *J. Chromatogr. A* 1292 (2013) 132–141, <https://doi.org/10.1016/j.chroma.2012.12.007>.
- [33] T. Gödecke, J.G. Napolitano, M.F. Rodríguez-Brasco, S.N. Chen, B.U. Jaki, D.C. Lankin, G.F. Pauli, Validation of a generic quantitative ¹H NMR method for natural products analysis, *Phytochem. Anal.* 24 (2013) 581–597, <https://doi.org/10.1002/pca.2436>.
- [34] W.A. Reed, L. Rao, P. Zanonato, A.Y. Garnov, B.A. Powell, K.L. Nash, Complexation of U VI with 1-Hydroxyethane-1,1-diphosphonic acid in acidic to basic solutions, *Inorg. Chem.* 46 (2007) 2870–2876, <https://doi.org/10.1021/ic0621401>.
- [35] B.M. Huang, S.Y. Xiao, T.B. Chen, Y. Xie, P. Luo, L. Liu, H. Zhou, Purity assessment of ginsenoside Rg1 using quantitative ¹H nuclear magnetic resonance, *J. Pharm. Biomed. Anal.* 139 (2017) 193–204, <https://doi.org/10.1016/j.jpba.2017.02.055>.
- [36] A. Kaur, R. Prakash, A. Ali, ¹H NMR assisted quantification of glycerol carbonate in the mixture of glycerol and glycerol carbonate, *Talanta* 178 (2018) 1001–1005, <https://doi.org/10.1016/j.talanta.2017.08.103>.
- [37] FDA, Guidelines for the Validation of Chemical Methods for the FDA Foods Program, *Foods Progr. Sci. Res. Steer. Comm.*, (2015), p. 37 <https://www.fda.gov/downloads/scienceresearch/fieldscience/ucm273418.pdf>.
- [38] FDA, Analytical Procedures and Methods Validation for Drugs and Biologics, *Guid. Ind.* (2015), pp. 1–15 http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1_Guideline.pdf.
- [39] S.L.R. Ellison, A. Williams, Eurachem working group on uncertainty in chemical measurement., quantifying uncertainty in analytical measurement, *Eurachem*, <https://www.eurachem.org/index.php/publications/guides/quam>, (2012), Accessed date: 23 February 2018.
- [40] M. López-García, R. Romero-González, M. Lacasaña, A. Garrido Frenich, Semiautomated determination of neonicotinoids and characteristic metabolite in urine samples using TurboFlow™ coupled to ultra high performance liquid chromatography coupled to Orbitrap analyzer, *J. Pharm. Biomed. Anal.* 146 (2017) 378–386, <https://doi.org/10.1016/j.jpba.2017.08.026>.
- [41] H. Chen, J. Marín-Sáez, R. Romero-González, A. Garrido Frenich, Simultaneous determination of atropine and scopolamine in buckwheat and related products using modified QuEChERS and liquid chromatography tandem mass spectrometry, *Food Chem.* 218 (2017) 173–180, <https://doi.org/10.1016/j.foodchem.2016.09.075>.
- [42] J. Ammon, et al., Guide to NMR method development and validation – Part 1: Identification and quantification, *EUROLAB Tech. Rep.* 1/2014, 2014 www.eurolab.org, Accessed date: 18 January 2018.
- [43] The European Commission, Sante/11813/2017 Analytical Quality Control and Method Validation Procedures for Pesticide Residues Analysis, (2018), pp. 1–46, <https://doi.org/10.13140/RG.2.2.33021.77283>.

CAPÍTULO TERCERO. APLICACIÓN DE qNMR A LA DETERMINACIÓN DE TÓXICOS EN AGUAS



4 **CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA**

4.1 **Introducción**

La investigación que recoge esta Memoria sobre la caracterización metabolómica de bebidas alcohólicas y de q NMR al control de la seguridad en aguas se ha dividido en dos secciones diferenciadas en función de los objetivos perseguidos. En la primera de ellas se estudia la aplicación de herramientas de la metabolómica en la clasificación e identificación de marcadores químicos de calidad en bebidas alcohólicas mediante GC-MS, LC-MS y NMR.

En cada uno de los estudios se han aplicado diferentes extracciones, así como diferentes técnicas de análisis en función de los analitos objetivo. La extracción de aquellos compuestos volátiles/semivolátiles del ron se ha llevado a cabo mediante HS-SPME para su análisis mediante GC-MS. Por otro lado, para aquellos componentes no volátiles, se ha empleado la extracción tipo “*dilute and shoot*” seguida de adquisición tipo LC-MS. El análisis de los rones mediante MS se ha llevado a cabo mediante *full scan*, obteniéndose así la obtención de la mayor cantidad de información de los analitos presentes.

En cuanto a la fracción mayoritaria del ron, ha sido analizada mediante NMR por análisis directo de los mismos tras adición de disolvente deuterado. Por último, se ha aplicado la técnica de fusión de datos.

La segunda parte de la presente Tesis recoge la investigación realizada sobre la aplicación de la NMR cuantitativa para la determinación de

compuestos tóxicos en aguas. A tal fin, se han desarrollado métodos de adquisición basados en núcleos magnéticamente activos de boro y fósforo respectivamente para la cuantificación de los analitos tanto mediante calibración interna como externa. En ambos casos, se han llevado a cabo tratamientos sencillos de preparación de las muestras basados en diluciones con agua deuterada.

La investigación realizada en la presente Tesis ha tenido como denominador común la aplicación de recursos analíticos establecidos como nuevas estrategias para afrontar retos comunes en el ámbito de la seguridad agroalimentaria. Los resultados obtenidos en cada una de las partes diferenciadas se discutirán a continuación.

4.2 Aplicación de herramientas de la metabolómica al estudio de ron

4.2.1 Adquisición de rones

En primer lugar, se llevó a cabo la adquisición de muestras de ron para el presente estudio metabolómico. El objetivo ha sido conseguir un conjunto de muestras representativas que aporten información suficiente para obtener modelos multivariantes óptimos. Para ello se seleccionaron distintas marcas de ron teniendo en cuenta las siguientes características:

- Bebidas alcohólicas categorizadas como ron en su etiquetado (Figura 9)
- Estado de conservación óptimo en el establecimiento (sin luz directa, temperatura ambiente, precinto intacto, etc.)
- Comercialmente disponibles y accesibles a cualquier consumidor
- Diversidad en la procedencia de elaboración



- Distintas edades de maduración entre la misma marca
- Maduración en barricas de madera (roble francés, americano, etc.)
- Destilación en distintos sistemas (alambique o columnas de cobre)
- Diferentes materias primas (azúcar de caña o melazas)

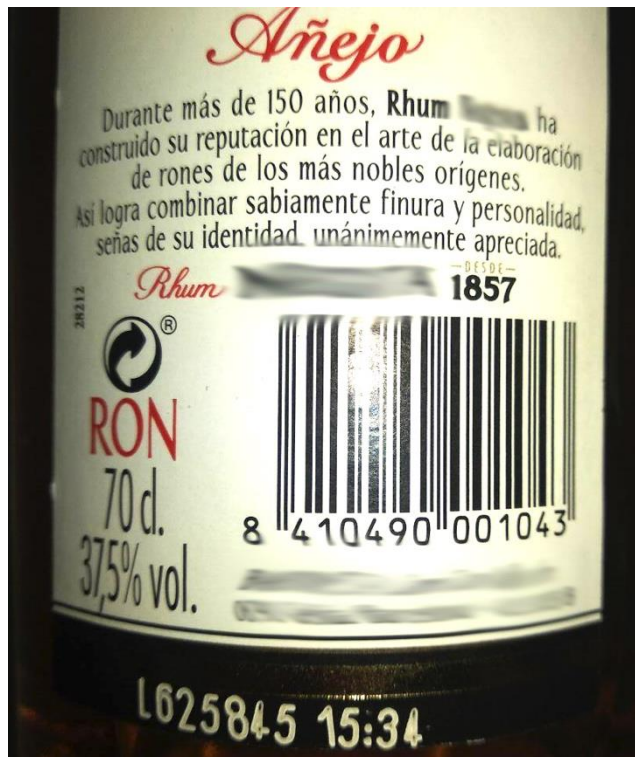


Figura 9. Ejemplo de etiquetado de un ron dorado comercial

Se adquirieron un total de 33 muestras de ron correspondientes a 16 marcas diferentes. Todas las muestras se almacenaron a temperatura ambiente antes del análisis en sus botellas de vidrio originales. La

información sobre la producción de ron, obtenida del sitio web oficial de los fabricantes de ron así como de la etiqueta y los códigos asignados para cada ron se resume en la Tabla 11. Aquella información sobre el envejecimiento, la materia prima y el barril que no era proporcionada por los fabricantes fue clasificada como N/A.

4.2.2 *Análisis de la fracción volátil*

La primera parte de la investigación que se recoge en este capítulo se dedicó a la optimización de un proceso de pretratamiento y de extracción de la fracción volátil/semivolátil de los rones, previo a su análisis mediante GC-QqQ-MS.

Se usaron fibras SPME de tipo polidimetilsiloxano (PDMS) de 100 μm , para la extracción de compuestos volátiles/semivolátiles con un rango de polaridad alta. La temperatura inicial del horno fue ajustada a 35°C, permitiendo así la elución de los compuestos más volátiles. Para evitar la coelución cromatográfica del etanol a alta concentración, se trabajó con una ventana de adquisición de 50–400 m/z (Figura 10). Se optimizaron parámetros como el volumen de muestra, tiempo de incubación, temperatura de extracción, tiempo de extracción y velocidad de agitación con el objetivo de obtener el mayor número analitos posible.



Tabla 11. Etiquetado de botellas e información del fabricante para las muestras de ron analizadas

Nombre	Origen	Edad	Materia prima	Método de destilación	Madera del barril
Havana 5 años	Cuba	5 años	Melazas	Columnas de acero inoxidable	Barril de roble francés
Havana 7 años		7 años	Melazas	Columnas de acero inoxidable	Barril de roble francés
Legendario Añejo		<i>Añejo</i>	Melazas	N/A	Barril de roble americano
Legendario 5 Años		Young	Melazas	N/A	Barril de roble americano
Legendario Elixir		N/A	Melazas	N/A	Barril de roble americano
Matusalem 7 Años	República Dominicana	7 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Matusalem 10 Años		10 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Matusalem 15 Años		15 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Almirante Añejo		<i>Añejo</i>	N/A	N/A	N/A
Almirante Miel		N/A	N/A	N/A	N/A
Baraguá Old Reserve	Grenada	Antigua reserva	N/A	N/A	N/A
Bohio Añejo		<i>Añejo</i>	N/A	N/A	N/A
Plantation	Trinidad & Tobago	Antigua reserva	Melazas	N/A	Mezcla (Barril de roble americano + French oak barrel)
Plantation Overproof		N/A	Melazas	Columnas de acero inoxidable	Mezcla (Barril de roble americano +

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

Nombre	Origen	Edad	Materia prima	Método de destilación	Madera del barril
					French oak barrel)
Botrán 8 Años	Guatemala	8 años	Azúcar de caña	Columnas de acero inoxidable	Barril de roble americano
Botrán 12 Años		12 años	Azúcar de caña	Columnas de acero inoxidable	Barril de roble americano
Botrán 18 Años		18 años	Azúcar de caña	Columnas de acero inoxidable	Barril de roble americano
Zacapa		6-25 años	Azúcar de caña	Columnas de acero inoxidable	Barril de roble americano
Appleton 5 Años	Jamaica	5-10 años	Melazas	Alambiques de cobre	Barril de roble americano
Appleton 7 Años		12 años	Melazas	Alambiques de cobre	Barril de roble americano
Flor de Caña 7 Años	Nicaragua	7 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Flor de Caña 12 Años		12 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Flor de Caña 18 Años		18 años	Melazas	Columnas de acero inoxidable	Barril de roble americano
Negríta Miel	España	N/A	N/A	N/A	N/A
Negríta		<i>Añejo</i>	Melazas	N/A	N/A
Artemi		<i>Dorado</i>	N/A	N/A	N/A
Tabay		<i>Añejo</i>	N/A	N/A	N/A
Reylik		N/A	N/A	N/A	N/A



Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabólica de bebidas alcohólicas y al análisis de tóxicos en aguas

Nombre	Origen	Edad	Materia prima	Método de destilación	Madera del barril
Santa Teresa 3 Años	Venezuela	3 años	Melazas	Mezcla (Columnas de acero inoxidable +Alambiques de cobre)	Barril de roble francés
Santa Teresa 1726		10 años	Melazas	Mezcla (Columnas de acero inoxidable +Alambiques de cobre)	Barril de roble francés
Velero		3 años	N/A	N/A	N/A

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

Adicionalmente se desarrolló un método de confirmación para el ron mediante GC-Q-Orbitrap. A tal fin, se usaron las mismas condiciones de extracción y análisis del método previamente descrito mediante GC-QqQ-MS. Las condiciones y parámetros del método desarrollado se resumen en la Tabla 12.

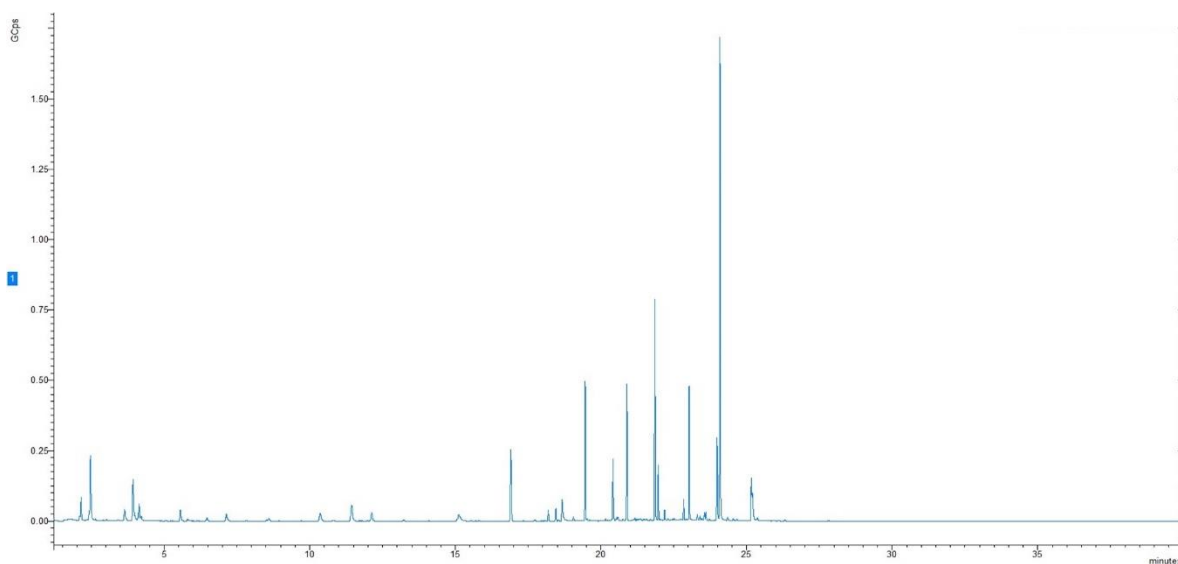


Figura 10. Cromatograma obtenido del análisis de una muestra de ron dorado mediante HS-SPME-GC-MS.



Tabla 12. Parámetros y condiciones para el análisis metabolómico de rones mediante HS-SPME-GC-MS

Condiciones HS-SPME-GC	
Volumen de muestra	20 mL
Tipo de fibra	100 μ m PDMS
T^a agitador	65 °C
Velocidad de agitación	250 r.p.m.
T^a de inyector	250 °C
Split	2 min (100:1), 9 min (20:1)
Columna	VF-5 ms (30 m x 0.25mm, d.i. 0.25 μ m)
Rampa de T^a	2 min a 35 °C, 100 °C a 4°C/min, 250 °C a 20°C/min, 20 min a 250 °C
Tiempo de análisis	43.75 min
Condiciones MS	
Modo de adquisición	<i>Full Scan</i>
Modo de ionización	EI
Rango de masas	50-400 m/z

4.2.3 Análisis de la fracción no volátil

Para el análisis de la fracción no volátil de los rones mediante LC-MS, se llevó a cabo un pretratamiento sencillo: 1 mL de cada ron fue diluido con 5 mL de agua Milli-Q y 10 μ L de la disolución resultante fueron directamente inyectados en el sistema.

De nuevo, las condiciones de separación y adquisición aplicadas se seleccionaron con la intención de obtener el máximo de información posible de un mismo análisis. A tal fin se aplicó una ventana de adquisición de hasta 1000 m/z para adquirir información espectral no sólo de metabolitos de baja masa molecular, sino también de otros analitos de mayor masa molecular esperables en bebidas, como azúcares. Las condiciones aplicadas se encuentran resumidas en la Tabla 13.

Tabla 13. Parámetros y condiciones para el análisis metabolómico de ronnes mediante LC-Orbitrap

Condiciones LC-Orbitrap	
Columna	Hypersil™ ODS C18
Volumen de inyección	10 µL
Fase móvil A	Acetato amónico (30 mM)
Fase móvil B	Metanol
Flujo	0.2 mL/min
Tiempo de análisis	58 min
Condiciones Orbitrap-MS	
Modo de adquisición	Alternativo: <i>Full Scan</i> , ESI+, ESI-
Resolución	25000 FWHM
Tiempo de scan	0.10 s
Rango de masas	50-1000 <i>m/z</i>

En la Figura 11 puede observarse un cromatograma típico obtenido de una muestra de ron dorado mediante LC-HRMS (Orbitrap-MS) en modo de ionización positiva y negativa.

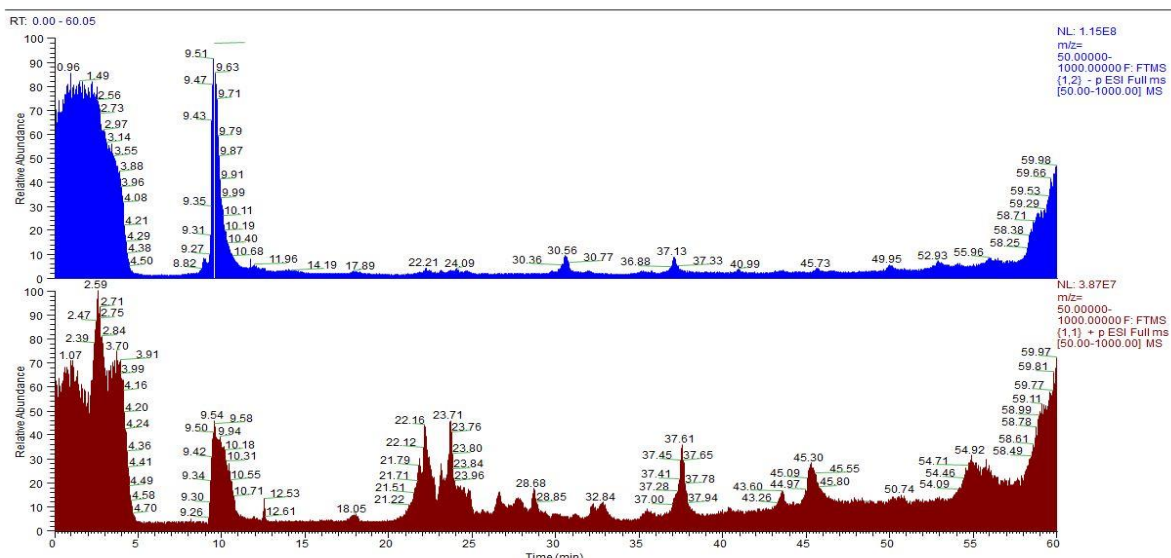


Figura 11. Cromatograma obtenido del análisis de ron dorado mediante LC-HRMS (Orbitrap-MS) en modo de ionización negativo (arriba) y positivo (abajo)



4.2.4 *Análisis mediante NMR*

Para el análisis de los componentes mayoritarios de los rones mediante NMR, a 0,5 mL de muestra se les añadió agua deuterada (100 μ L) con TMSP como estándar interno de referencia (0.05 %). Los análisis fueron realizados en modo de adquisición de protón ($^1\text{H-NMR}$) en un equipo de alta resolución de 600 MHz equipado con criosonda.

La Tabla 14 recoge los parámetros y valores usados en la adquisición de rones dorados mediante NMR para su análisis metabolómico.

Tabla 14. Parámetros y condiciones para el análisis metabolómico de rones mediante NMR

Condiciones de adquisición de NMR	
Núcleo observado	Hidrógeno
Frecuencia de adquisición	600.130 MHz
Disolvente	D ₂ O
T^a de adquisición	27°C
Estándar interno	TMSP
Tiempo de relajación	2.4 segundos
Tiempo de adquisición	2.7 segundos
Nº de scans	32

La supresión de señales con pulsos selectivos a distintas frecuencias para la eliminación de las mismas se llevó a cabo con el objetivo de evitar las distorsiones generadas por el agua y etanol presente en los rones. Además, se realizaron análisis con desacoplamiento de carbono (^{13}C), evaluando así la posibilidad de futuros análisis multivariantes con señales sin presencia de satélites procedentes de la actividad magnética

del ^{14}C . La Figura 12 muestra un ejemplo de los espectros obtenidos de un ron dorado en los tres distintos modos de adquisición estudiados: a) sin supresión de señales ni desacoplamiento de carbono, b) con supresión de señales y sin desacoplamiento de carbono y c) con supresión de señales y desacoplamiento de carbono.

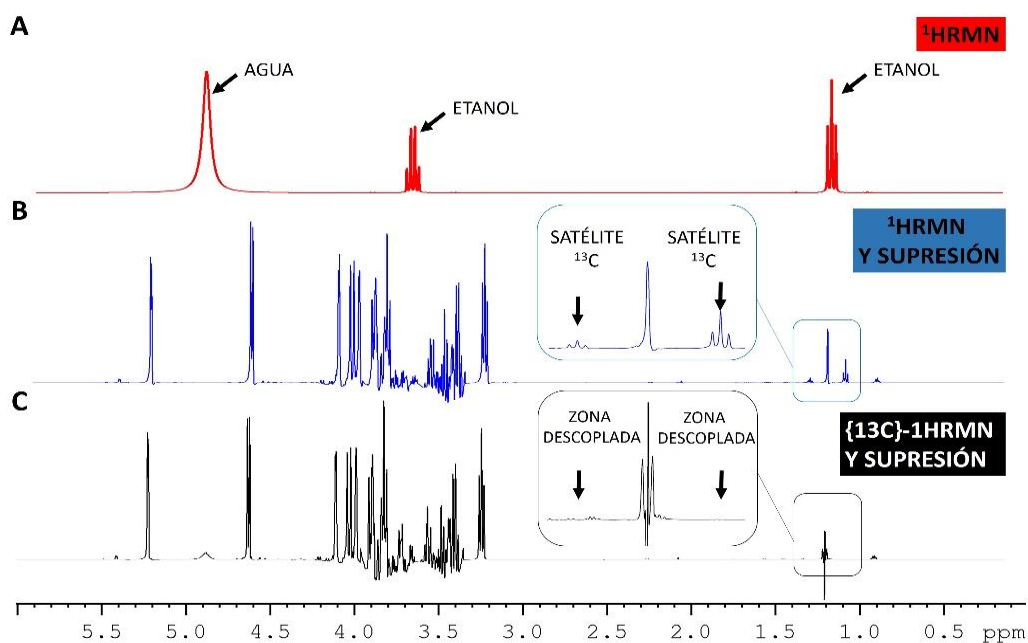


Figura 12. Espectro de NMR de un ron dorado entre 6.0-0.0 ppm: a) sin supresión ni desacoplamiento de señales, b) con supresión de señales y sin desacoplamiento de carbono y c) con supresión de señales y desacoplamiento de carbono



4.2.5 Estudio metabolómico de rones

Para los datos obtenidos mediante GC-MS, se hizo uso del software *MZmine*. El pretratamiento de dichos datos ha consistido en los siguientes procesos: a) deconvolución de picos, b) alineamiento de señales, c) generación de una lista de picos, d) escalado, e) filtrado de duplicados y f) rellenado de ceros. Se estudiaron 232 variables iniciales para los 33 rones evaluados junto a sus réplicas. Por último, los datos se exportaron a una hoja de datos para su posterior procesado multivariante mediante la herramienta informática *SPSS*.

Con el objetivo de realizar un análisis exploratorio de los datos obtenidos se realizó primeramente un análisis HCA (Figura 13). Como resultado se detectaron un total de 4 *outliers* por diferencias notorias respecto a sus réplicas. Además, 5 muestras de rones fueron preliminarmente clasificadas dadas sus condiciones especiales de ser rones con aditivos como miel, sirope o condimentos (saborizantes de vainilla y mandarina).

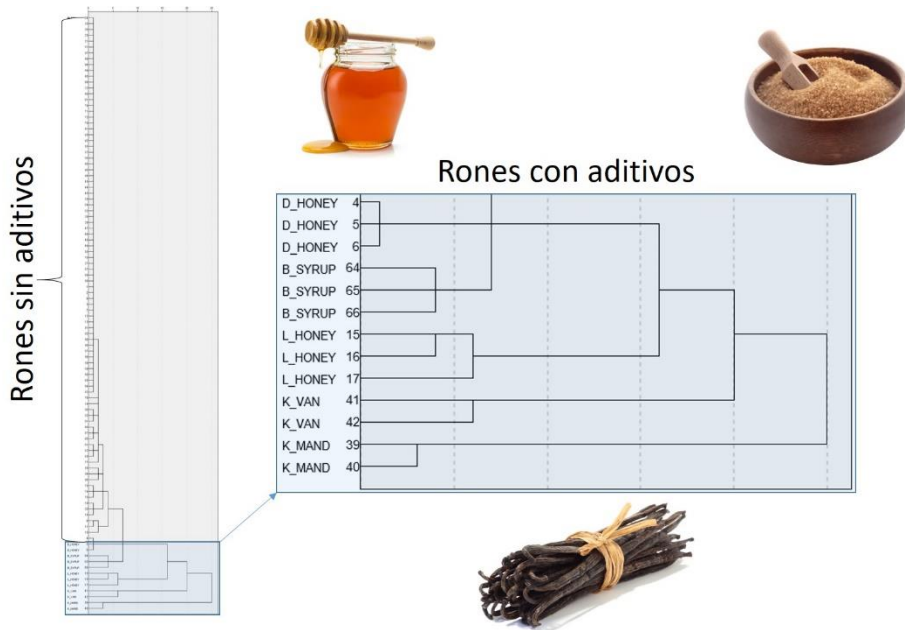


Figura 13. Dendrograma obtenido mediante HCA del conjunto total de rones

Por otro lado, se aplicó el PCA, seleccionando las variables con un mayor índice de poder discriminante aportado por el *software*. Un total de 40 iones fueron elegidos, correspondientes a 13 compuestos diferentes.

El PCA obtenido (Figura 14) explicó el 88% de la variación total. El PC1, PC2 y PC3 explicaron el 36%, 36% y 15% de la varianza, respectivamente. En dicho análisis, no se observó ningún patrón ni agrupación de los rones destacado pese a la alta varianza obtenida, remarcando la complejidad de las múltiples características de los procesos de elaboración (envejecimiento, tipo de barril, método de destilación, materia prima, etc.). A fin de obtener modelos clasificatorios de las distintas categorías propuestas, se llevaron a cabo análisis supervisados.

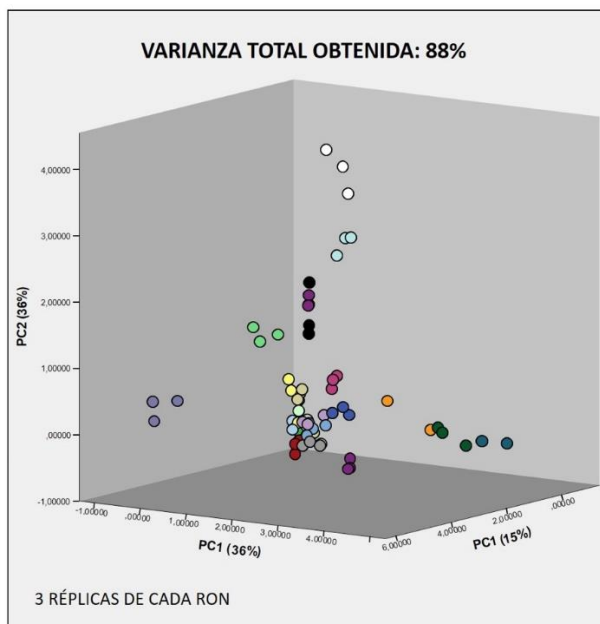


Figura 14. Representación PCA del conjunto de rones basado en aquellas variables más discriminantes

Para llevar a cabo los análisis LDA, el método de exclusión permutada de uno a uno (método *leave-one-out*) se utilizó como procedimiento de validación cruzada.

Los resultados mostraron excelentes resultados de clasificación mediante modelos LDA para las distintas categorías seleccionadas: origen ($R^2=0.94$ y $Q^2=0.73$), materia prima ($R^2=0.97$ y $Q^2=0.91$), método de destilación ($R^2=0.92$ y $Q^2=0.87$), barril ($R^2=0.95$ y $Q^2=0.86$) y edad ($R^2=0.94$ y $Q^2=0.65$). En la Figura 15 se muestran las representaciones gráficas de los LDA obtenidos para cada una de las clasificaciones estudiadas.

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

Los resultados obtenidos demostraron la alta capacidad clasificatoria de los modelos desarrollados, con valores obtenidos para R^2 entre 0.94 y 0.97, así como valores entre 0.91 y 0.65 para Q^2 . Aquellos rones cuya información no es aportada en los etiquetados (rones N/A) pudieron ser tentativamente clasificados. El parámetro de envejecimiento del ron se estudió considerando los rones de menos de 5 años como jóvenes, los de entre 5-10 años como medios y los rones de mayor edad como rones viejos.

En cuanto a la materia prima, los resultados de LDA revelaron que los rones de los que se carecía de información podrían estar elaborados a partir de melaza. Se observó una clara diferenciación entre los rones en columnas de acero inoxidable respecto a alambique de cobre. Para las clasificaciones acerca del tipo de barril utilizado, la proximidad en el gráfico LDA de los barriles de roble sin especificar y roble americano podría ser una indicación de que el roble americano es la madera típica utilizada cuando los fabricantes no especifican el tipo de barril. En relación con la edad LDA permite una buena diferenciación de rones jóvenes, medios y viejos (Figura 15).

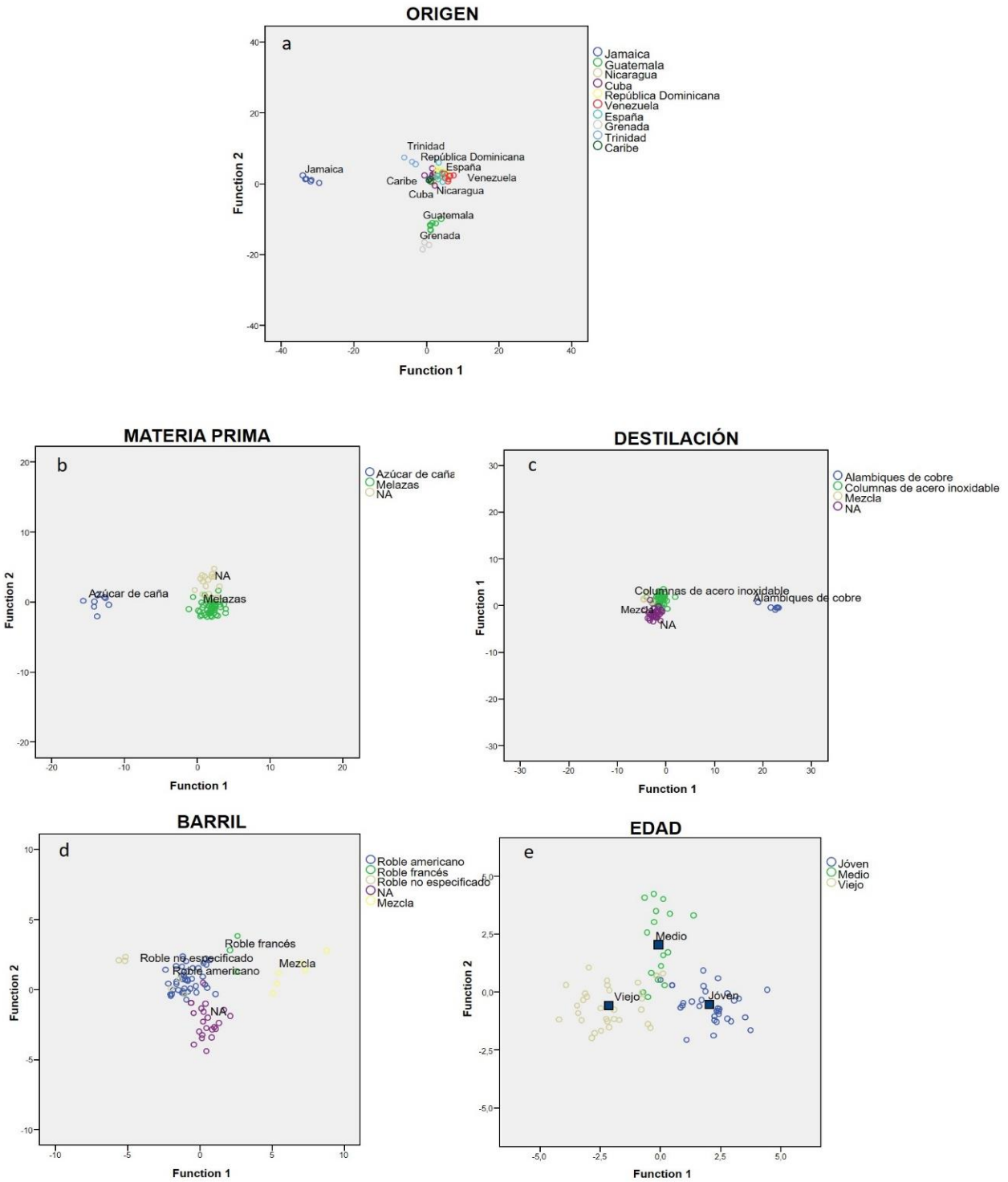


Figura 15. Representación LDA del conjunto de roncs basado en a) Origen, b) Materia prima, c) Destilación, d) Barril y e) Edad

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

Para la identificación de aquellas variables más discriminantes del estudio, el análisis mediante HS-SPME-GC-MS y HS-SPME-Q-Exactive-GC permitió una asignación tentativa de los analitos en base a diferentes factores: comparación mediante librería NIST de alta y baja resolución, índice de retención de Kovats (KI) y clasificaciones de puntuación HRF (filtrado de alta resolución) aportado por el *software*.

De los compuestos tentativamente identificados, seis de ellos fueron identificados como etil ésteres (Tabla 15): ácido etil éster hexadecanoico, ácido etil éster (E)-9-octadecanoico, ácido etil éster tetradecanoico, ácido 3-metilbutil éster octadecanoico, ácido etil éster decanoico y ácido etil éster octanoico. Éste tipo de componentes son comunes como productos de fermentación de bebidas alcohólicas con un papel importante en las propiedades organolépticas de las bebidas fermentadas, debido a su aroma afrutado.

Tabla 15. *Compuestos tentativamente identificados como más discriminantes en el estudio metabolómico de roncs mediante HS-SPME-GC-MS y datos obtenidos para su identificación*

Compuesto	Fórmula empírica	<i>m/z</i>	RT (min)	KI ^a	HRF ^b (%)	Método de identificación
Ácido etil éster hexadecanoico	C ₁₈ H ₃₆ O ₂	185.2	24.10	1993	100.00	KI. Lib. HRF
		241.2				
		284.3				
		213.2				
		157.2				
		115.1				
		88.1				
55.1						
Ácido etil éster (E)-9-octadecanoico	C ₂₀ H ₃₈ O ₂	180.2	25.17	2170	100.00	KI. Lib. HRF
		112.2				
		222.3				
		264.2				
		137.1				
		109.1				



Compuesto	Fórmula empírica	<i>m/z</i>	RT (min)	KI ^a	HRF ^b (%)	Método de identificación
		169.1				
Ácido etil éster tetradecanoico	C ₁₆ H ₃₂ O ₂	88.1 55.1	23.05	1789	100.00	KI. Lib. HRF
Ácido 3-metilbutil éster octadecanoico	C ₁₃ H ₂₆ O ₂	70.2 127.2	20.89	1441	99.83	KI. Lib. HRF
Ácido etil éster decanoico	C ₁₂ H ₂₄ O ₂	157.1 115.1 88.1 55.1	20.44	1389	99.50	KI. Lib. HRF
Ácido etil éster octanoico	C ₁₀ H ₂₀ O ₂	88.1 127.2 57.2	18.17	1194	99.86	KI. Lib. HRF
Ianona	C ₁₃ H ₁₈	131.1 159.1 144.1 116.1 91.1 113.1	18.4	1211	100.00	KI. Lib. HRF
1,1-dietoxi-3-metil butanona	C ₉ H ₂₀ O ₂	103.1 73.1 75.1	9.2	918	99.82	KI. HRF
Derivado de la ianona	C ₁₃ H ₁₈	159.1	19.5	1295	99.53	HRF
Desconocido	-	130.2	18.2	1197	-	-
Acetato de etilo	C ₄ H ₈ O ₂	54.7	2.3	611	100.00	KI. Lib. HRF
Tetrahidro 2-piran-2-metanol	C ₆ H ₁₂ O ₂	85.1	11.8	988	100.00	KI. Lib. HRF
1,1-dietoximetano	C ₅ H ₁₂ O ₂	103.1	3.3	668	-	KI. Lib

KI; Índice de retención de Kovats

HRF; Valor de filtrado de alta resolución (*High-Resolution Filtering score*)

Asimismo, se identificaron tentativamente otros componentes como acetato de etilo, 1,1-dietoxi-3-metil butanona o 1,1-dietoximetano, con menor poder de discriminación. Este tipo de compuestos derivados de la fermentación alcohólica tienen una relación estrecha con la preparación de los rones, siendo comúnmente relacionados con el sabor y olor de los mismos. Tan sólo un analito a un tiempo de retención de 18,2 minutos e ion 130,2 m/z no fue identificado debido a la ausencia de candidatos estructurales en las librerías disponibles.

La Tabla 15 muestra los 13 compuestos más discriminantes del estudio multivariante junto a sus tiempos de retención, masas detectadas, valor de KI, valor HRF (%) y nombre correspondiente para aquellos identificados.

Adicionalmente, se llevó a cabo una monitorización de los ésteres etílicos detectados. Para ello, se estudiaron seis rones de marcas diferentes y de distintas edades centrándose en la intensidad de las señales detectadas para ésteres etílicos tipo C8 a C16. Los resultados confirmaron una relación directa entre el envejecimiento del ron y la presencia de dichos componentes en ellos, tal como puede observarse en la Figura 16.

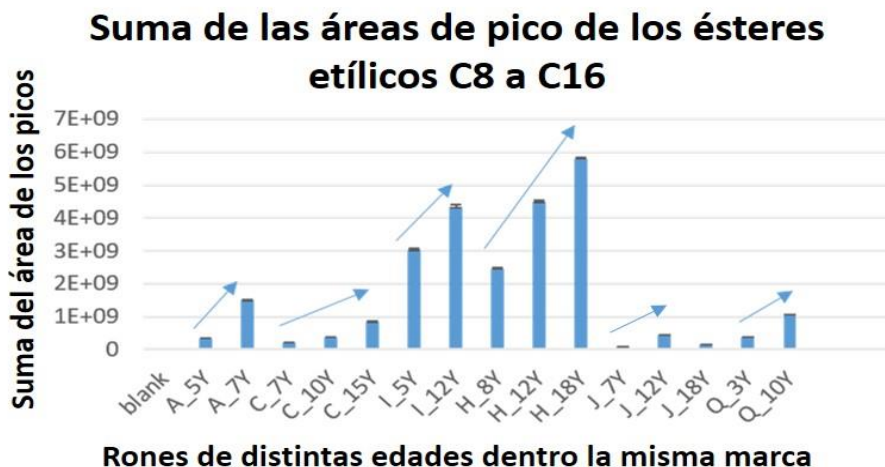


Figura 16. Representación de barras de la suma de etil ésteres en rones e indicaciones de tendencias mediante flechas

Respecto al estudio mediante LC-MS, para llevar a cabo el pretratamiento de los datos previamente obtenidos se usó el software *Compound Discoverer*. A dicho fin, se llevó a cabo la a) selección de picos, b) integración de las señales, c) alineamiento de los tiempos de retención, d) substracción de blancos y e) exportación a hoja de datos, para los siguientes tratamientos multivariantes mediante el *software SIMCA*. En nuestro caso, se seleccionaron un total de 1020 variables. Se aplicó la transformación logarítmica a todos los compuestos detectados, dando valores en un rango más pequeño sin enmascarar el efecto de valores pequeños dentro de los datos. Además, se llevó a cabo un escalado tipo Pareto previo a los análisis multivariantes para reducir la influencia de los picos intensos y enfatizar los picos menos intensos que pudieran tener mayor relevancia metabólica.

El análisis mediante LC-MS de los rones proporcionó un número ampliamente mayor de variables respecto a las obtenidas mediante GC-

MS. Éste hecho se debió a dos razones principales: la menor selectividad de la técnica de extracción de *dilute-and-shoot* frente a HS-SPME, así como al mayor número de compuestos no volátiles presentes en la composición del ron.

El análisis no supervisado de los datos obtenidos se llevó a cabo mediante PCA para obtener una visualización general de las relaciones entre las muestras del modelo (Figura 17). Se detectaron y eliminaron un total de 5 *outliers* de los análisis multivariantes (Figura 18). Los resultados mostraron cómo un total del 82% de varianza fue explicada para el modelo propuesto para un total de 11 PCs.

A continuación, se crearon diferentes modelos PLS-DA para cada uno de los grupos de clasificación: origen, envejecimiento, materia prima y barril. El método de destilación no fue incluido en el estudio mediante LC-MS debido a que los resultados obtenidos no mostraron clasificaciones apropiadas, tal y como sí sucedía con el estudio de la fracción volátil.

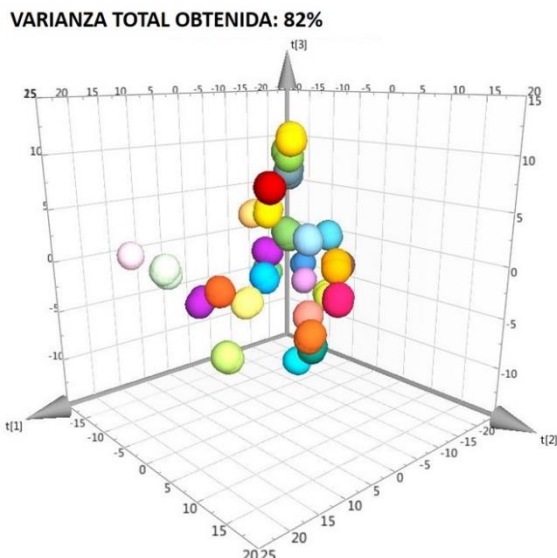


Figura 17. Representación tridimensional del PCA obtenido para el conjunto de rones mediante LC-MS

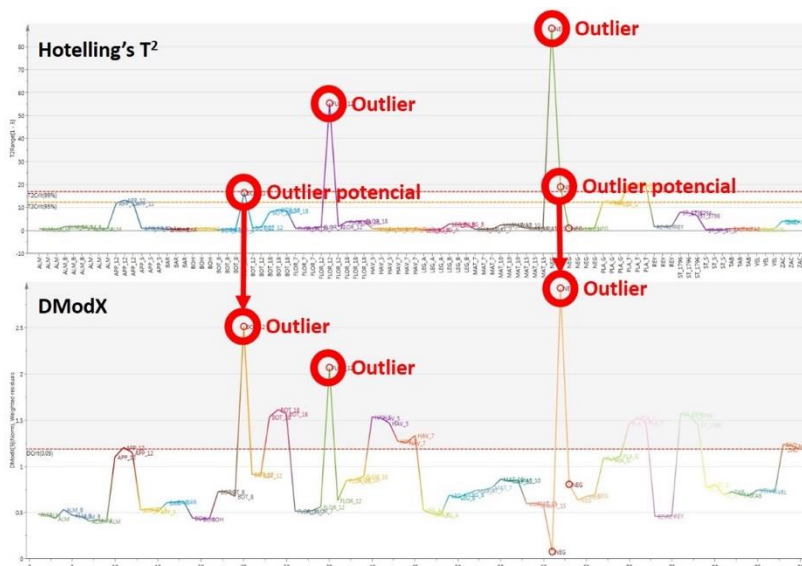


Figura 18. Representación gráfica de Hotelling's T^2 (arriba) y DmodX (abajo) junto a los outliers detectados (límites de confianza= 95-99%)

El parámetro de envejecimiento del ron para PLS-DA se estudió considerando los rones con una edad igual o inferior a 7 años como rones

jóvenes y envejecidos durante más de 7 años, a diferencia del estudio mediante GC-MS donde la edad del ron se clasificó como joven, medio o viejo. De ésta forma se delimitaron las posibilidades de clasificación respecto a la edad, dificultadas por el proceso opcional de mezclado de rones de distintas edades con el que cuenta su preparación. Además, se trató de mejorar la representatividad del conjunto de muestras añadiendo tres rones de distinta marca considerados como blancos y, por ende, no añejados previamente en barricas de madera. Aquellos rones con aditivos estudiados previamente fueron descartados debido a las grandes diferencias que mostraron respecto al resto de rones, siendo el número final de muestras analizadas de 30 junto a sus réplicas.

Como mejora para la validación de los modelos en los análisis supervisados se introdujo la división al azar del total de muestras en dos conjuntos. El primero de ellos, conjunto de entrenamiento, quedó conformado por un 4/5 de las muestras, mientras el 1/5 de las muestras restantes se reservaron como conjunto de validación externa (conjunto de predicción). De esta forma se llevó a cabo una evaluación de la fiabilidad de los resultados de forma externa, obteniendo valores de CC% entre 94 -100% para cada modelo propuesto.

Los resultados obtenidos para los diferentes modelos PLS-DA realizados en base a las distintas categorías seleccionadas junto a los valores de R^2 y Q^2 se muestran en la Figura 19. Como pudo observarse, el análisis supervisado de los rones en base al origen de elaboración, pese a tratarse de un modelo intencionadamente sobreajustado en busca de tendencias, no aportó ninguna correlación destacable ($R^2Y= 0.774$ y $Q^2Y= 0.600$).



Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabolómica de bebidas alcohólicas y al análisis de tóxicos en aguas

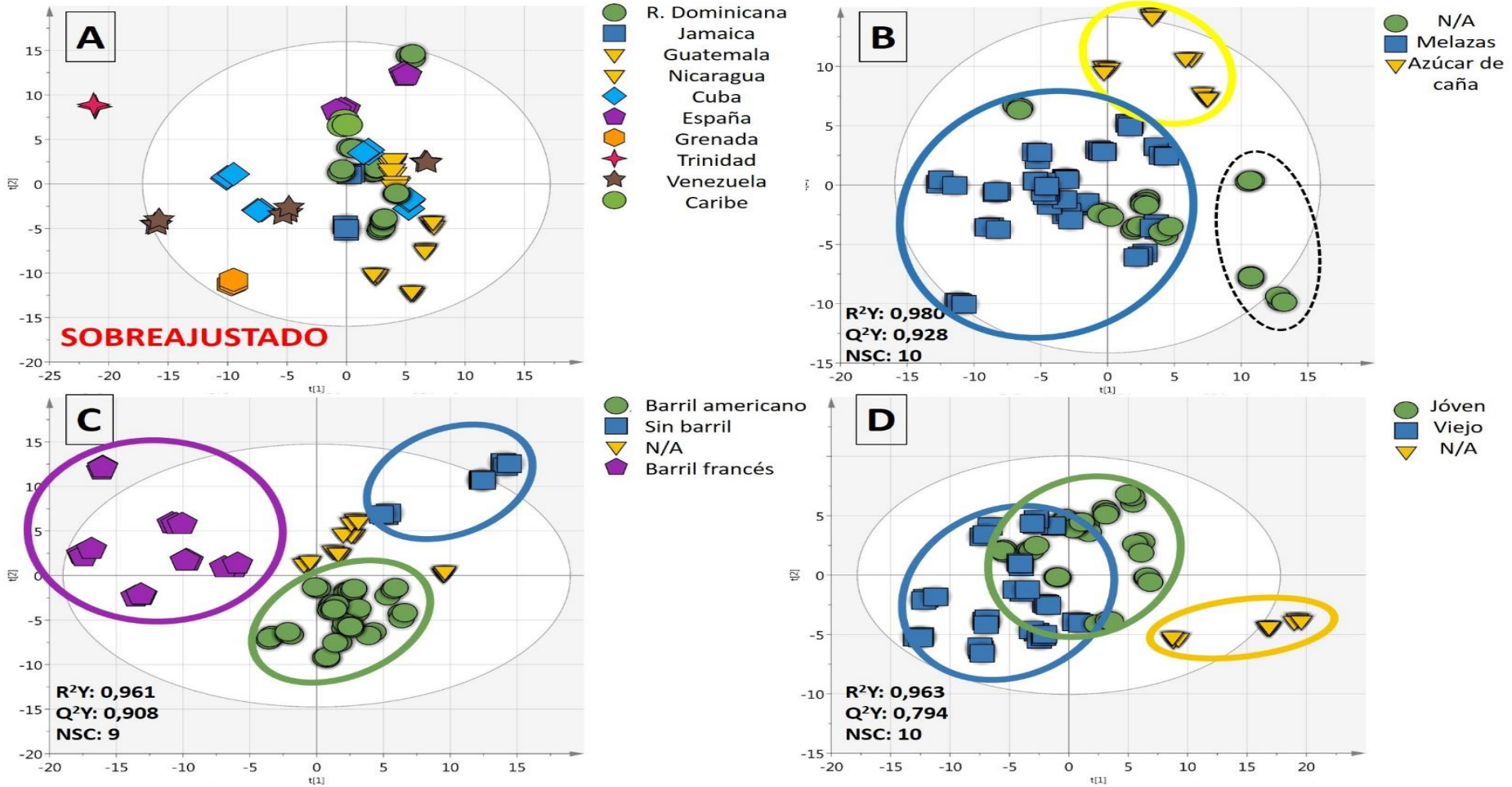


Figura 19. Representación PLS-DA basada en LC-MS del conjunto de roncs basado en: a) Origen, b) Materia prima, c) Barril y d) Edad

Aquellos rones para los cuales se desconocía el uso de melaza o caña de azúcar, se clasificaron mayormente dentro del grupo de melaza, sugiriendo su uso durante el proceso de fermentación. Los tres rones blancos adicionales se diferenciaron del resto de los rones, lo que sugirió el uso de materias primas diferentes o alternativas para su preparación.

Tal como sucedía en el estudio metabolómico mediante GC-MS, las diferencias entre el tipo de barril fueron concluyentes. La distribución gráfica entre el conjunto de muestras indicó que la madera de roble americano parece usarse cuando los fabricantes no especifican el tipo de barril utilizado. Además, para la clasificación por edades se observó una tendencia lineal en la representación gráfica que depende del envejecimiento del ron. Un menor valor de Q^2 obtenido para las clasificaciones por edad de los rones puede explicarse por las diferencias en la legislación sobre el etiquetado, donde la composición de las mezclas no es fidedigna a la reflejada en el mismo.

La obtención de los espectros de masas en los modos de ionización positivo y negativo dio lugar a la posibilidad de estudiar la identidad de los analitos en ambos modos. A tal fin, se evaluó el error de masa, el ajuste isotópico, la evaluación pseudo MS/MS y la coincidencia con bases de datos. Cuando se encontraron varios candidatos para un m/z dado, la identificación del metabolito se clasificó dando preferencia a los que normalmente se encuentran en las bebidas alcohólicas. De nuevo, la alta resolución, incrementó la capacidad de identificación de los metabolitos con mayor capacidad discriminatoria en cada uno de los estudios mediante PLS-DA. El análisis VIP (valor de filtrado >1.5) de los distintos modelos supervisados propuestos permitió la selección de



10 de ellos. Además, la alta concentración de otros 6 analitos propició su selección para ser identificados adicionalmente (Tabla 16).

Entre los compuestos identificados tentativamente, se encontraron diferentes derivados de azúcar como los compuestos más discriminantes. Esto se justifica dado que tanto la melaza como el azúcar de caña contienen altos valores de sacáridos como sacarosa, glucosa y fructosa, los cuales tras el proceso de fermentación y maduración dan lugar a dichos derivados. Además, la presencia de furfurales se encuentra directamente relacionada con la composición de los rones mediante la descomposición de la celulosa de las barricas en el proceso de maduración. Los resultados confirmaron un fuerte efecto de diferentes productos de azúcar como la fructosilglicina (VIP 2.04) y 1- deoxy-1-nitro-D-manitol (VIP 1.66) y de furfurales como el 5-hidroxiacetilfurfural (VIP 1.79) en las categorías seleccionadas.

De igual modo, se detectaron otros derivados del azúcar como 1-desoxi-1-nitro-D-manitol, metil 4,6-didesoxi-4- (3-desoxi-L-glicero-tetronamida)-2-O-metil- α - D-manopiranosida y la metil-6-desoxi- α -D-glucopiranosida, con alta capacidad discriminante para la clasificación de rones. Compuestos con menor capacidad discriminante, pero presentes en alta concentración, como la hexosa y la galactosamina, confirmaron la manifiesta importancia de los azúcares en el estudio metabolómico de la fase no volátil de los rones.

Tabla 16. Compuestos identificados tentativamente como más discriminantes en el conjunto de rones mediante LC-Orbitrap-MS

RT (min)	Nombre	Fórmula	Aducto	m/z Teórico	m/z Experimental	Δ masa (ppm)	Fragmento MS/MS	VIP
9.199	Fructosilglicina	C ₈ H ₁₅ NO ₇	[M+H] ⁺	237.08485	237.08483	0.09	133.04954 115.02639	2.04 (Barril)
9.001	1-Deoxi-1-nitro-D-manitol	C ₆ H ₁₃ NO ₇	[M+H] ⁺	211.06920	211.06938	0.84	-	1.71 (Edad) 1.66 (Materia prima)
10.726	L-Prolina	C ₅ H ₉ N O ₂	[M+H] ⁺ [M+Na] ⁺	115.06333	115.06411	6.76	70.06513	1.83 (Materia prima) 1.63 (Edad)
11.797	Ácido 4-Guanidinobutanoico	C ₅ H ₁₁ N ₃ O ₂	[M+H] ⁺	145.08513	145.08532	1.31	-	1.98 (Materia prima)
25.311	5,6-dihidro-1H-Ciclopenta[c]furan-1,3(4H)-diona	C ₇ H ₈ O ₄	[M+H] ⁺	156.04225	156.04237	0.73	115.03897 129.01824	1.52 (Materia prima) 1.50 (Barril)



Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabolómica de bebidas alcohólicas y al análisis de tóxicos en aguas

RT (min)	Nombre	Fórmula	Aducto	m/z Teórico	m/z Experimental	Δmasa (ppm)	Fragmento MS/MS	VIP
27.520	Ácido dietil ester L-(+)-Tartárico	C ₈ H ₁₄ O ₆	[M+Na] ⁺ [M+H] ⁺	206.07904	206.07906	0.09	-	1.91 (Materia prima)
35.911	Metil 4,6-dideoxi-4-(3-deoxi-L-glicero-tetronamido)-2-O-metil-α-D-manopiranosida	C ₁₂ H ₂₃ NO ₇	[M+H] ⁺	293.14746	293.14752	0.21	133.04954 115.03897	1.58 (Materia prima)
9.200	Ácido(-)-Treo-isodihomocitríco	C ₈ H ₁₂ O ₇	[M+Na] ⁺ [M+H] ⁺ [M+NH ₄] ⁺	220.05830	220.05817	0.61	157.04954	1.57 (Barril)
16.473	Metil 6-deoxi-alfa-D-glucopiranosida	C ₇ H ₁₄ O ₅	[M+NH ₄] ⁺	178.08412	178.08435	1.30	73.02841	1.51 (Barril)
23.077	Hidroximetilfurfural	C ₆ H ₆ O ₃	[M+Na] ⁺ [M+H] ⁺ [M+NH ₄] ⁺	126.03169	126.03201	2.51	81.03349 109.02841	1.98 (Edad) 1.79 (Barril)
9.718	Hexosa	C ₆ H ₁₂ O ₆	[M+NH ₄] ⁺	180.06338	180.06277	3.43	85.02933 145.04954	<1.5
9.770	Galactosamina	C ₆ H ₁₃ NO ₅	[M+H] ⁺	179.07938	179.07947	0.52	127.03897 145.04954	<1.5

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

RT (min)	Nombre	Fórmula	Aducto	m/z Teórico	m/z Experimental	Δ masa (ppm)	Fragmento MS/MS	VIP
9.846	Ácido D-Gluconico	C ₆ H ₁₂ O ₇	[M-H] ⁻	196.05830	196.05719	4.68	89.02442	<1.5
30.065	Acetato-(+)-Mentol	C ₁₂ H ₂₂ O ₂	[M-H] ⁻	198.16199	198.16125	3.70	197.15361	<1.5
30.066	Homovanillil alcohol	C ₉ H ₁₂ O ₃	[M+H] ⁺	168.07864	168.07875	0.65	109.06479 154.06245	<1.5
30.197	Ácido 5-[(4-Metoxifenoxi)metil]-2-furoico	C ₁₃ H ₁₂ O ₅	[M+H] ⁺	248.06847	248.06868	0.86	81.03349 109.02841	<1.5



Asimismo, se identificaron tentativamente otros analitos directamente relacionados con las bebidas alcohólicas como agentes potenciadores de sabor: derivados de la vanillina, como el alcohol homovanillil, o del mentol, como el mentol acetato. La Tabla 16 muestra los 16 compuestos más discriminantes del estudio multivariante junto a sus tiempos de retención, fórmula, aductos detectados, peso teórico y experimental, error de masa, fragmentos detectados y valor de VIP asignado.

4.2.6 Fusión y comparación de datos multi-técnica

El conjunto de ronnes fue también analizado mediante NMR y como último paso se fusionaron los resultados obtenidos con los tres conjuntos de datos (GC-MS, LC-MS y NMR). Los datos espectrales obtenidos mediante NMR de 600 MHz fueron pretratados a través de la herramienta *NMRProcFlow* previo a su tratamiento multivariante. Para ello, se llevó a cabo el alineamiento de las señales respecto al estándar interno de referencia (TMSP, 0.00 ppm), supresión de las regiones correspondientes a las resonancias del etanol (1.17 y 3.65 ppm) y agua (4.76 ppm), así como la segmentación en porciones de 0.1 ppm. De ésta forma, se obtuvieron un total de 100 segmentos espectrales para cada una de las muestras de ron analizadas.

Como software para la aplicación de herramientas estadísticas se utilizó *SIMCA*. El procesado aplicado a los datos fue: transformación matemática logarítmica y escalado tipo Pareto previo a los análisis multivariantes. De igual forma, se realizó una evaluación de la fiabilidad de los resultados de forma externa, obteniendo valores de CC% de 100% para cada modelo propuesto. Debido a la cualidad que presentan las

adquisiciones mediante NMR, donde distintos grupos de resonancias pueden pertenecer a un componente individual, la evaluación inicial de la varianza de los resultados mediante análisis no supervisado fue fundamental.

En primer lugar, se compararon los resultados con dos modos de adquisición diferentes: ^1H -NMR acoplada y desacoplada con ^{13}C . A tal fin, se evaluó el rendimiento de clasificación de los modelos PLS-DA a través de los valores de $R^2\text{X}$, $R^2\text{Y}$ y $Q^2\text{Y}$ obtenidos. Aunque se obtuvieron buenos resultados para los modelos $^1\text{H} \{^{13}\text{C}\}$ - NMR, se obtuvieron mejores resultados para los modelos ^1H -NMR, tal como puede observarse en la Tabla 17. Estos resultados pueden explicarse en términos de degradación de la resolución espectral de las señales observadas de ^1H -NMR debido a las supresiones, que pueden dar lugar a pérdidas de información discriminante crucial en dichas regiones.

Tabla 17. Comparación de los resultados de PLS-DA obtenidos para los distintos tipos de adquisición mediante NMR

Adquisición ^1H -NMR				
Parámetro	Edad	Materia prima	Barril	Método de destilación
$R^2\text{Y}$	0.934	0.927	0.870	0.747
$Q^2\text{Y}$	0.824	0.851	0.803	0.669
NSC	6	7	5	7
Adquisición $^1\text{H}\{^{13}\text{C}\}$ -NMR				
Parámetro	Edad	Materia prima	Barril	Método de destilación
$R^2\text{Y}$	0.911	0.920	0.835	0.690
$Q^2\text{Y}$	0.774	0.872	0.794	0.639
NSC	6	6	5	7

Los resultados preliminares obtenidos mediante ^1H -NMR unidimensional permitieron obtener una huella dactilar que corroboró la alta variabilidad de los rones, tal como sucedía mediante las técnicas



anteriores y quedó patente en el PCA ($R^2= 89\%$) que se muestra en la Figura 20. Pese a que dicho modelo no supervisado no contó con información previa de las agrupaciones de las muestras del análisis, tras una inspección visual pudo deducirse una clara tendencia en términos de edad.

Tras un planteamiento de 4/5 de las muestras como conjunto de entrenamiento y 1/5 conjunto de predicción para una validación externa, se llevó a cabo el análisis multivariante supervisado de los resultados obtenidos. La Figura 21 muestra los modelos PLS-DA obtenidos mediante adquisición $^1\text{H-NMR}$ para el conjunto de rones basado en a) edad, b) barril, c) destilación y e) materia prima. Tal como sucediera en el análisis metabolómico mediante LC-MS, el análisis supervisado de los rones en base al origen de elaboración no mostró resultados firmes.

Los resultados obtenidos aportaron información acerca de las condiciones de preparación para aquellas muestras donde no se obtuvo información en el etiquetado: uso de barricas de roble americano para el envejecimiento y una mayor proximidad a los grupos de melazas así como también el uso de destilaciones mediante acero inoxidable. Estos dos últimos parámetros mostraron una mayor diferenciación en las representaciones bidimensionales (Figura 21) respecto a los demás grupos, sugiriendo así ciertas diferencias de preparación respecto a los métodos convencionales. Finalmente, la clasificación por edad, tal como pudo verse en el análisis no supervisado previo, mostró una separación clara en términos de rones jóvenes y viejos.

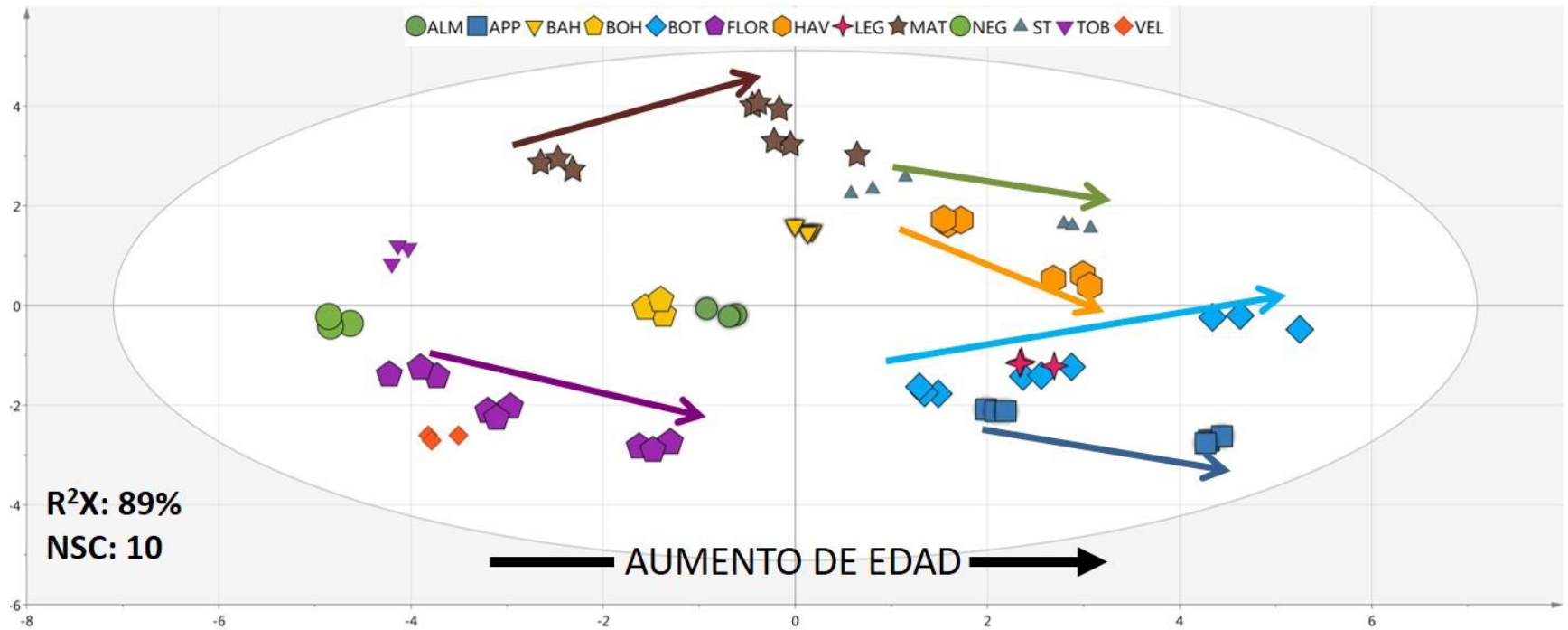


Figura 20. Representación del PCA obtenido del análisis del conjunto de rones mediante NMR





Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabolómica de bebidas alcohólicas y al análisis de tóxicos en aguas

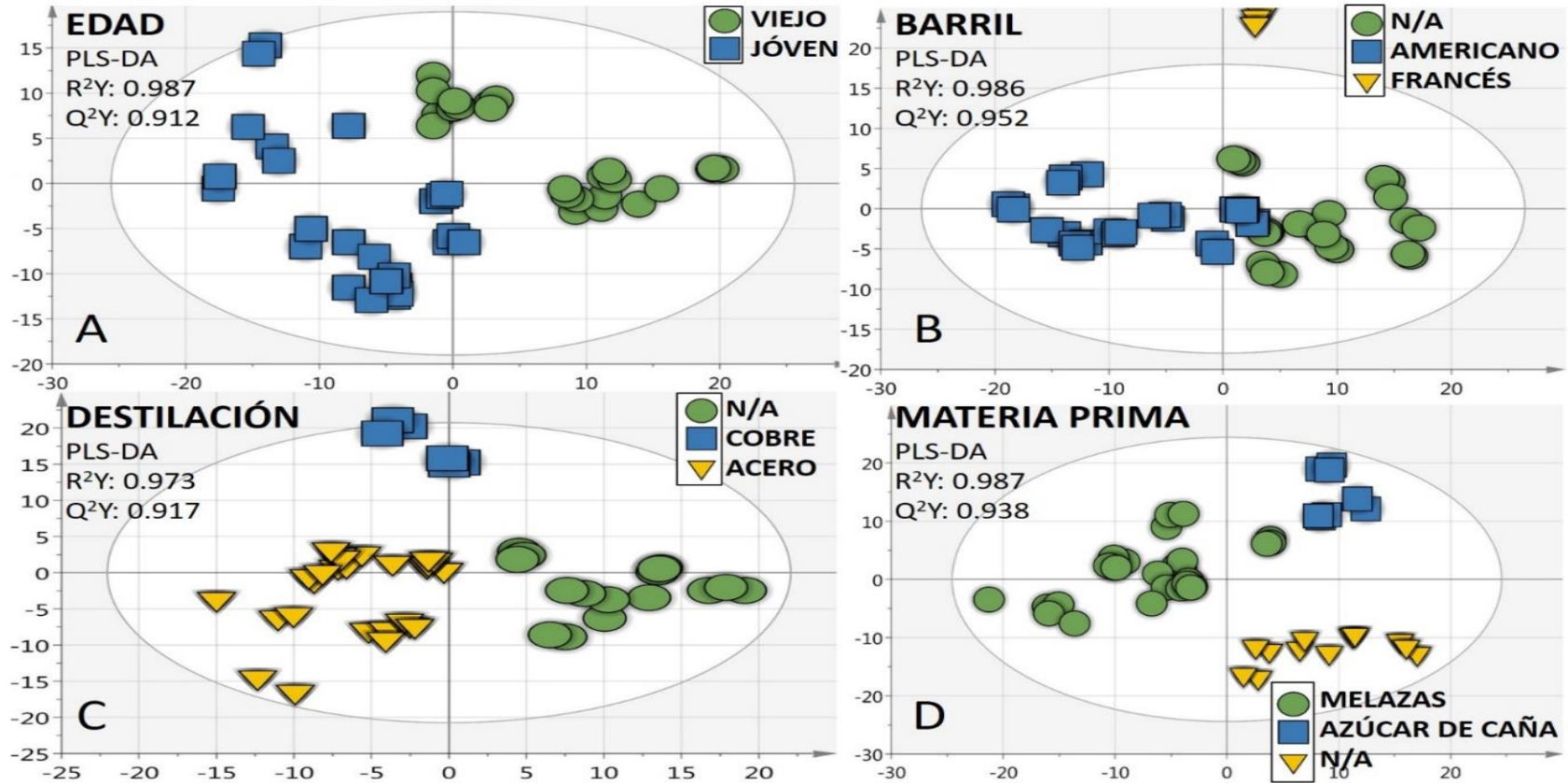


Figura 21. Representación PLS-DA basada en 1H -NMR del conjunto de rones basado en: a) Edad, b) Barril, c) Destilación y d) Materia prima

A continuación, con la información conjunta de las técnicas de NMR, GC-MS y LC-MS, se procedió a la fusión de datos. Tanto en la fusión de nivel bajo como medio, se trabajó la combinación de bloques de datos para obtener luego un modelo con una interpretación conjunta mejorada. A tal fin, la fusión de nivel bajo se realizó mediante la fusión de los datos “crudos”, generando un bloque de datos global que incluyó la información sin tratamiento previo obtenida por las tres técnicas. Para la fusión de nivel medio, se trabajó con las variables resultantes, en éste caso las más discriminantes (mediante análisis VIP), de la reducción de dimensionalidad de los bloques de datos individuales ya tratados, fusionándose e interpretándose luego conjuntamente.

El nivel bajo de fusión de datos demostró ser la aproximación más apropiada para la clasificación de los roncs, tal como puede observarse en la Tabla 18. En ella, se muestran los resultados obtenidos individualmente para cada una de las distintas técnicas, así como en la fusión baja y media aplicada. El conjunto total de datos de todas y cada una de las técnicas contó con 100 variables para NMR, 232 variables para GC-MS y 1020 variables para LC.

Para el caso de la fusión de datos de nivel medio, la selección de variables se basó en análisis VIP con filtros de discriminación de $VIP > 1.5$ para LC-MS y $VIP > 1.0$ para GC-MS y $^1\text{H-NMR}$. El valor VIP mayor para LC-MS fue debido al mayor número de variables obtenidas en la adquisición, consiguiendo de ésta forma un volumen razonable de datos multi-técnica. Finalmente, la fusión de datos de bajo nivel demostró proporcionar los mejores resultados en comparación con las técnicas individuales, consiguiendo el 100% de muestras clasificadas correctamente para todas las categorías estudiadas.



Tabla 18. Valores de validación obtenidos por PLS-DA para el análisis metabolómico de las distintas técnicas individualmente y mediante fusión de datos de nivel bajo y medio

¹H-NMR				
Parámetro	Edad	Materia prima	Barril	Método de destilación
R²Y	0.934	0.927	0.870	0.747
Q²Y	0.824	0.851	0.803	0.669
NSC	6	7	5	7
GC-MS				
Parámetro	Edad	Materia prima	Barril	Método de destilación
R²Y	0.916	0.971	0.957	0.939
Q²Y	0.712	0.843	0.739	0.747
NSC	5	8	10	7
LC-MS				
Parámetro	Edad	Materia prima	Barril	Método de destilación
R²Y	0.928	0.977	0.977	0.975
Q²Y	0.765	0.918	0.922	0.910
NSC	3	6	6	6
FUSIÓN DE DATOS DE NIVEL BAJO				
Parámetro	Edad	Materia prima	Barril	Método de destilación
R²Y	0.987	0.987	0.986	0.973
Q²Y	0.912	0.938	0.952	0.917
NSC	5	6	7	6
FUSIÓN DE DATOS DE NIVEL MEDIO				
Parámetro	Edad	Materia prima	Barril	Método de destilación
R²Y	0.882	0.902	0.897	0.888
Q²Y	0.779	0.822	0.850	0.814
NSC	3	4	4	4
Variables	50	27	38	43

Tal como puede observarse en la Figura 22, los resultados obtenidos para los modelos PLS-DA mediante fusión de datos de nivel bajo dieron lugar a las clasificaciones más claras mediante representaciones bidimensionales. Los resultados descritos previamente acerca del uso de

barricas de roble, melazas y destilación en alambiques de acero inoxidable fueron confirmados para aquellos rones cuya información inicial no era completa. Además, en la clasificación mediante PLS-DA en base a la edad de los rones se puede encontrar indicios de una posible región correspondiente a rones de mediana edad (7-8 años).

4.3 Aplicación de la qNMR a la determinación de compuestos tóxicos

4.3.1 Determinación de ácido bórico en biocidas comerciales

La qNMR (^{11}B -NMR) ha sido aplicada a la determinación de ácido bórico (Figura 7) en formulados de biocidas que contienen esta sustancia como procedimiento de control de calidad de producto. Los análisis de NMR se realizaron en un espectrómetro de 300 MHz equipado con una sonda BBFOPLUS (^1H , BB- ^{19}F , *SmartProbe*). Los espectros obtenidos fueron ajustados a desplazamientos de 0.00 ppm usando BF_3OEt_2 como patrón estándar de referencia. Las condiciones de detección optimizadas se encuentran resumidas en la Tabla 19.

Aplicación de la espectrometría de masas y la resonancia magnética nuclear a la caracterización metabolómica de bebidas alcohólicas y al análisis de tóxicos en aguas

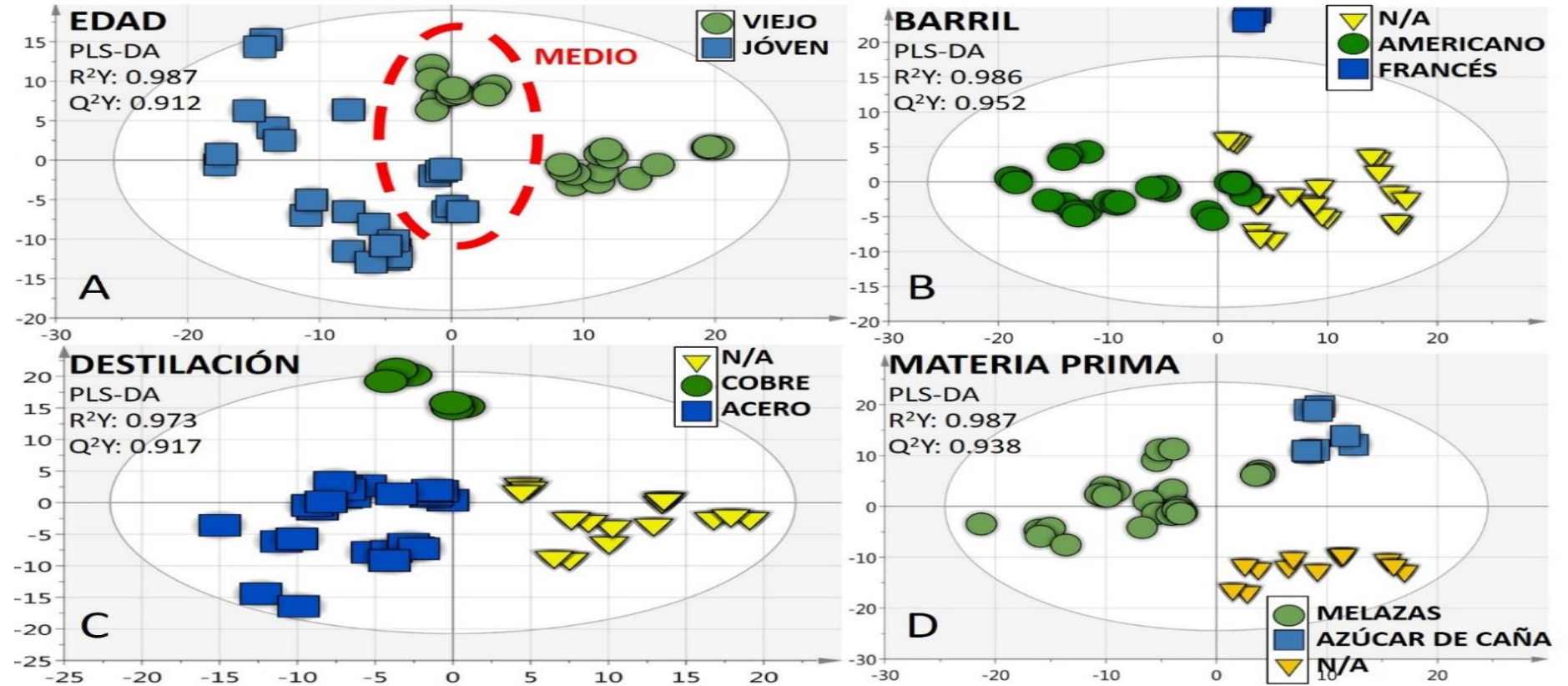


Figura 22. Representación PLS-DA basada en fusión de datos de nivel bajo del conjunto de roncs basado en: a) Edad, b) Barril, c) Destilación y d) Materia prima

Tabla 19. Parámetros y condiciones para la determinación de ácido bórico mediante NMR

Condiciones de adquisición de NMR	
Núcleo observado	Boro
Tubos	Cuarzo (5mm)
T^a de adquisición	20 °C
Desplazamiento químico	19.4 ppm
Ventana	199.7 ppm
Frecuencia de adquisición	96.3 Hz
Ángulo de pulso	90°
Disolvente	D ₂ O
Resolución espectral	1.565 Hz
Estándar interno	BF ₃ OEt ₂
Tiempo de relajación	0.1 segundos
Tiempo de adquisición	0.6 segundos
N° de scans	64

Uno de los principales problemas para la adquisición del núcleo de boro-11 mediante NMR fue la presencia del mismo en tubos de adquisición fabricados con borosilicatos. Éste hecho da lugar a la presencia de señales interferentes en torno a $\delta_B = -6.4$ ppm. Para evitarlo, se usaron tubos cuya composición mayoritaria era el cuarzo (contienen menos del 0.1% de boro).

Se optimizaron distintos parámetros de adquisición mediante el análisis de muestras de agua fortificadas con ácido bórico, como el número de *scans*, el tiempo de adquisición y el tiempo de relajación. Tal como se puede observar en la Figura 23, la correlación lineal encontrada dentro



del número de escaneos demostró que la reproducibilidad y precisión obtenida fue óptima para el estudio.

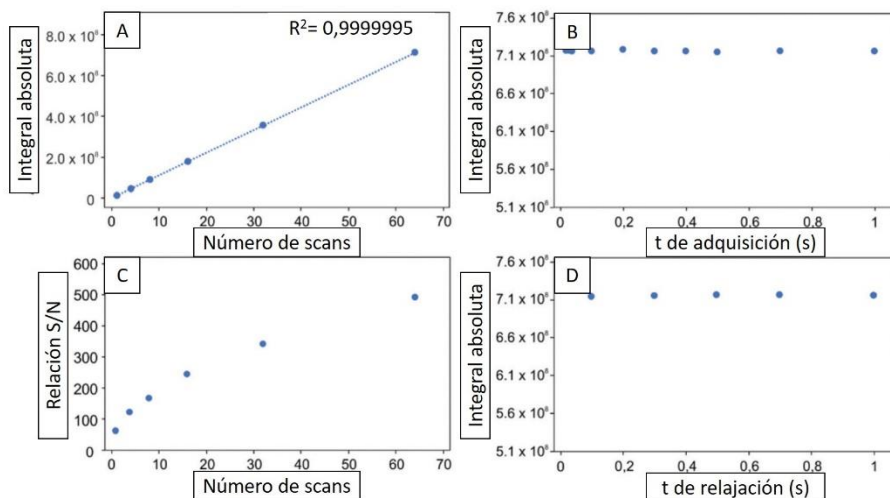


Figura 23. Influencia de: a) Número de scans en la señal absoluta, b) Número de scans en la relación S/N, c) Tiempo de adquisición y d) Tiempo de relajación en la señal absoluta

Por otro lado, el rango de trabajo del método de análisis propuesto fue evaluado con concentraciones entre 0.05% y 4.10% (el máximo valor de solubilidad de ácido bórico en agua), obteniéndose una linealidad de $R^2 = 0.9996$. La Figura 24 muestra los resultados de linealidad obtenidos en la recta de calibrado.

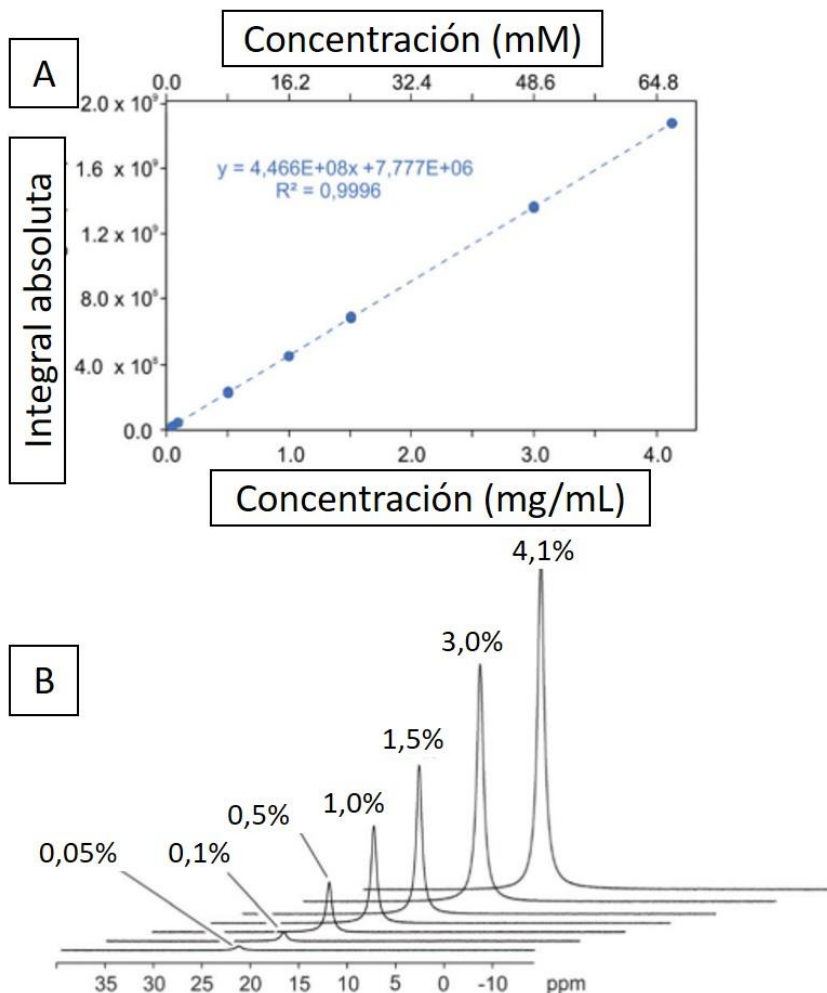


Figura 24. a) Linealidad obtenida para el ácido bórico en agua y b) Espectros de ^{11}B -NMR a lo largo del rango de trabajo

El método fue validado mediante establecimiento de distintos parámetros como el LOD, LOQ, recuperación, precisión intradía, precisión interdía e incertidumbre a distintos niveles de concentración (0.5 % p/p, 1.5 % p/p y 4.0% p/p). La Tabla 20 muestra resumidos los valores obtenidos.



Tabla 20. Parámetros de validación obtenidos para el método de ^{11}B -NMR propuesto.

Parámetro	0.5% (p/p)	1.5% (p/p)	4.0% (p/p)
Recuperación (%)	106.2	93.6	101.0
Precisión intradía (n=5) (%)	0.9	1.2	0.7
Precisión interdía (n=5) (%)	2.0	1.5	1
Incertidumbre (%)	4.4	3.9	3.7
LOD (%)	0.02		
LOQ (%)	0.04		
Linealidad	$R^2 = 0.999$		

Finalmente, se analizaron un total de 5 muestras de formulados comerciales de biocidas en pastillas que contenían ácido bórico para control de aguas de piscinas y embalses agrícolas. La preparación de las muestras consistió en la adición de 150 mg de las mismas en 1 mL de D_2O , seguido de agitación durante 10 minutos, centrifugación durante otros diez minutos adicionales, y análisis de 0.5 mL del sobrenadante resultante.

Las concentraciones encontradas variaron significativamente entre los diferentes tipos de productos investigados, obteniendo concentraciones desde $<\text{LOQ}$ al 10% (p/p). Los valores de concentración obtenidos se compararon con los de la técnica de ICP-MS como método externo de referencia, encontrando diferencias siempre menores que el 6% respecto a los obtenidos mediante ^{11}B -NMR.

4.3.2 Determinación de ácido etidróico en aguas de lavado de vegetales

Asimismo se ha empleado la q NMR para la cuantificación de HEDP (Figura 8), aditivo antioxidante adicionado a aguas para el lavado de frutas y hortalizas. Con el objetivo de desarrollar el método analítico, se optimizaron diversos parámetros de adquisición mediante ^{31}P - q NMR. Las condiciones de detección optimizadas para el método de adquisición de HEDP en agua mediante NMR se encuentran resumidas en la Tabla 21.

Tabla 21. Parámetros y condiciones para la determinación de ácido etidróico mediante NMR

Condiciones de adquisición de NMR	
Núcleo observado	Fósforo
Tubos	Borosilicato (5mm)
Tª de adquisición	20 °C
Desplazamiento químico	20.1 ppm
Ventana	395.7 ppm
Frecuencia de adquisición	121.5 Hz
Ángulo de pulso	90°
Disolvente	D ₂ O
Resolución espectral	1.467 Hz
Estándar interno	Na ₂ HPO ₄
Tiempo de relajación	0.05 segundos
Tiempo de adquisición	0.68 segundos
Nº de scans	116

Inicialmente, se trató de desarrollar una estrategia de cuantificación basada en la calibración interna, haciendo uso de un IS del núcleo observado (Na₂HPO₄, δ_{P} 3.1 ppm). Se llevó a cabo un análisis de los tiempos de relajación necesarios para asegurar una óptima relajación de los núcleos de fósforo-31 tanto del HEDP como del IS. Como puede





observarse en la Figura 25, se requería un mínimo $D1 = 10$ s para HEDP y un mínimo de $D1=60$ s para el IS, obteniendo así una relajación óptima. Debido a los resultados obtenidos y a la ausencia de patrones primarios con tiempos de relajación menores, se optó por llevar a cabo el desarrollo del método en base a una calibración externa, al igual que sucedió en el estudio del ácido bórico.

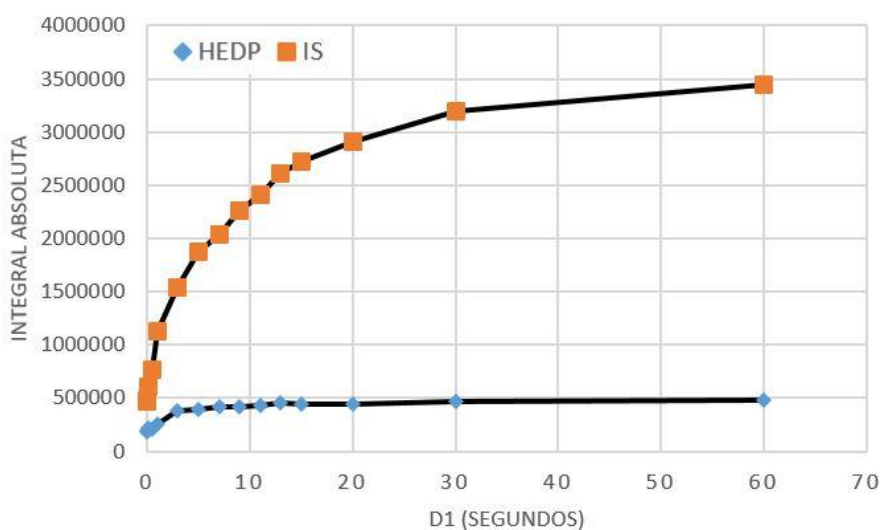


Figura 25. Determinación del tiempo de relajación óptimo mediante las señales de fósforo-31 del ácido etidróico y del IS

Se optimizó un valor óptimo entre el número de *scans* y la intensidad de la señal obtenida. La Figura 26 muestra cómo se obtiene una adecuada linealidad a medida que se aumenta el número de adquisiciones para una muestra de agua fortificada al 0.02% con HEDP.

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA

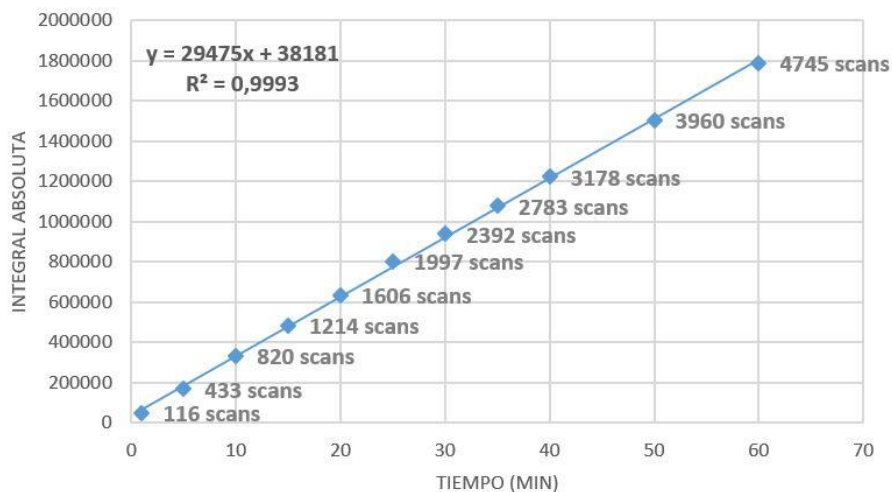


Figura 26. Influencia del número de scans sobre la integral absoluta del ácido etidróico

El método propuesto fue validado mediante el estudio de su selectividad, rango de trabajo, precisión, LOQ, estabilidad, veracidad e incertidumbre. Los parámetros de validación obtenidos para el método desarrollado de análisis de ácido etidróico mediante ^{31}P -qNMR se encuentran resumidos en la Tabla 22.

Tabla 22. Parámetros de validación obtenidos para el método de ^{31}P -NMR propuesto.

Parámetro	0.01% (p/p)	0.02% (p/p)	0.05% (p/p)
Recuperación (%)	99.9	-	96.7
Precisión intradía (n=5) (%)	1.58	0.99	0.88
Precisión interdía (n=5) (%)	4.30	2.22	1.19
Incertidumbre (%)	8.9	-	3.7
LOQ (%)	0.01 (116 scans) – 0.003 (4745 scans)		
Linealidad	$R^2 = 0.997$		



El método desarrollado fue aplicado a 10 muestras reales de aguas previamente utilizadas en industrias para el lavado de vegetales. Se detectó HEDP en dos de las muestras estudiadas a una concentración de 0.027% y 0.008%, respectivamente. El método analítico propuesto presentó una clara ventaja en relación otros métodos utilizados en términos de simplicidad, velocidad y sensibilidad.

CAPÍTULO CUARTO. DISCUSIÓN INTEGRADA





5 CAPÍTULO QUINTO. CONCLUSIONES

En la presente Tesis se ha hecho uso de recursos analíticos para el desarrollo de nuevas estrategias para afrontar retos comunes en el ámbito de la seguridad agroalimentaria. A tal fin se han desarrollado estudios metabolómicos para la compresión y clasificación de ron mediante GC-MS, LC-MS y NMR. Por otro lado, se han desarrollado, validado y aplicado métodos analíticos basados en *q*NMR para la determinación de ácido bórico y ácido etidróico en biocidas comerciales y aguas de lavado de vegetales, respectivamente.

En base a los resultados obtenidos en los estudios experimentales llevados a cabo, se destacan las siguientes conclusiones:

- Los resultados obtenidos mediante análisis *fingerprinting* de compuestos volátiles/semivolátiles, no volátiles y mayoritarios a través de GC-MS, LC-MS y NMR han demostrado un excelente comportamiento en las distintas categorías seleccionadas, asegurando resultados fiables y validados: $R^2Y > 0.90$ y $Q^2Y > 0.70$ en todos los casos para la clasificación de materias primas, tipo de barril y método de destilación. Aquellos rones cuya información inicial no era completa han sido asociados al uso de barricas de roble, melazas y destilación en alambiques de acero inoxidable. Además, se han determinado un total de 29 compuestos como los más discriminantes para las clasificaciones propuestas, destacando entre ellos los ácidos de ésteres etílicos, derivados azucarados y furfurales. La simpleza de los protocolos de tratamiento de muestra (HS-SPME, *dilute and shoot* y

disolución en disolventes deuterados) han supuesto una reducción de tiempo y costes importante, minimizando además errores humanos y contaminaciones.

- La fusión de datos multi-técnica (GC-MS, LC-MS y NMR), realizada por primera vez con las técnicas descritas en bebidas, ha dado lugar a resultados de clasificación significativamente mejores que con las técnicas individuales. De ésta forma se han obtenido R^2Y superiores a 0,97 para las categorías propuestas y se obtuvo con éxito un valor de clasificación del 100% de CC% para todas las muestras analizadas. La estrategia de fusión de datos de bajo nivel ha aportado los mejores resultados, indicando que la pérdida de información en las reducciones de dimensionalidad puede empeorar los resultados de un modelo.
- Los dos métodos desarrollados y validados para la determinación de analitos con núcleos magnéticamente activos de boro y fósforo mediante la aplicación de la técnica de q NMR han sido aplicados con excelentes resultados al análisis de muestras. En concreto, la cuantificación mediante ^{11}B -NMR supone un modo de trabajo sin precedentes en compuestos borados a través de q NMR. Las validaciones realizadas (selectividad, rango de trabajo, precisión interdía e intradía, LOQ, recuperación e incertidumbre) han demostrado la total aptitud de la técnica como alternativa a métodos más tediosos ampliamente extendidos hoy en día.



6 CAPÍTULO SEXTO. COPYRIGHT

Aportación científica 1. Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

J.R. Belmonte-Sánchez, S. Gherghel, J. Arrebola-Liébanas, R.R. González, J.L.M. Vidal, I. Parkin, A.G. Frenich. *Talanta*. 187 (2018) 348–356.



RightsLink®

Home

Account
Info

Help



Title: Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry

Author: José Raúl Belmonte-Sánchez, Simona Gherghel, Javier Arrebola-Liébanas, Roberto Romero González, José Luis Martínez Vidal, Ivan Parkin, Antonia Garrido Frenich

Publication: *Talanta*

Publisher: Elsevier

Date: 1 September 2018

© 2018 Elsevier B.V. All rights reserved.

Logged in as:
José Raúl Belmonte

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

Copyright © 2019 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Aportación científica 2. An Innovative Metabolomic Approach for Golden Rum Classification Combining Ultrahigh-Performance Liquid Chromatography–Orbitrap Mass Spectrometry and Chemometric Strategies

J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido French. *J. Agric. Food Chem.* 67 (2019) 1302–1311.



RightsLink®

Home

Account Info

Help



ACS Publications
Most Trusted. Most Cited. Most Read.

Title:

An Innovative Metabolomic Approach for Golden Rum Classification Combining Ultrahigh-Performance Liquid Chromatography–Orbitrap Mass Spectrometry and Chemometric Strategies

Author:

José Raúl Belmonte-Sánchez, Roberto Romero-González, Francisco Javier Arrebola, et al

Publication:

Journal of Agricultural and Food Chemistry

Publisher:

American Chemical Society

Date:

Jan 1, 2019

Copyright © 2019, American Chemical Society

Logged in as:

José Raúl Belmonte

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK

CLOSE WINDOW

Copyright © 2019 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customer@copyright.com



Aportación científica 3. $^1\text{H-NMR}$ and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis

J.R. Belmonte-Sánchez, L.M. Aguilera-Sáez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich. *Microchem. J.* 150 (2019).



RightsLink®



Home



Help



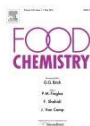
Email Support



Sign In



Create Account



1H-NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis

Author: José Raúl Belmonte-Sánchez, Roberto Romero-González, José Luis Martínez Vidal, Francisco Javier Arrebola, Antonia Garrido Frenich

Publication: Food Chemistry

Publisher: Elsevier

Date: 1 July 2020

© 2020 Elsevier Ltd. All rights reserved.

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW





Aportación científica 4. Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides

L.M. Aguilera-Sáez, J.R. Belmonte-Sánchez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich, I. Fernández. *Analyst*. 143 (2018).

Issue 19, 2018



From the journal:
Analyst

Pushing the frontiers: boron-11 NMR as a method for quantitative boron analysis and its application to determine boric acid in commercial biocides

L. M. Aguilera-Sáez, J. R. Belmonte-Sánchez, R. Romero-González, J. L. Martínez Vidal, F. J. Arrebola, A. Garrido Frenich and I. Fernández,
Analyst, 2018, **143**, 4707

DOI: 10.1039/C8AN00505B

If you are not the author of this article and you wish to reproduce material from it in a third party non-RSC publication you must [formally request permission](#) using Copyright Clearance Center. Go to our [Instructions for using Copyright Clearance Center page](#) for details.

Authors contributing to RSC publications (journal articles, books or book chapters) do not need to formally request permission to reproduce material contained in this article provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
Reproduced from Ref. XX with permission from the Centre National de la Recherche Scientifique (CNRS) and The Royal Society of Chemistry.



- o For reproduction of material from PCCP:
Reproduced from Ref. XX with permission from the PCCP Owner Societies.
- o For reproduction of material from PPS:
Reproduced from Ref. XX with permission from the European Society for Photobiology, the European Photochemistry Association, and The Royal Society of Chemistry.
- o For reproduction of material from all other RSC journals and books:
Reproduced from Ref. XX with permission from The Royal Society of Chemistry.

If the material has been adapted instead of reproduced from the original RSC publication "Reproduced from" can be substituted with "Adapted from".

In all cases the Ref. XX is the XXth reference in the list of references.

If you are the author of this article you do not need to formally request permission to reproduce figures, diagrams etc. contained in this article in third party publications or in a thesis or dissertation provided that the correct acknowledgement is given with the reproduced material.

Reproduced material should be attributed as follows:

- For reproduction of material from NJC:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the Centre National de la Recherche Scientifique (CNRS) and the RSC
- For reproduction of material from PCCP:
[Original citation] - Reproduced by permission of the PCCP Owner Societies
- For reproduction of material from PPS:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry (RSC) on behalf of the European Society for Photobiology, the European Photochemistry Association, and RSC
- For reproduction of material from all other RSC journals:
[Original citation] - Reproduced by permission of The Royal Society of Chemistry

If you are the author of this article you still need to obtain permission to reproduce the whole article in a third party publication with the exception of reproduction of the whole article in a thesis or dissertation.

Information about reproducing material from RSC articles with different licences is available on our [Permission Requests page](#).



Aportación científica 5. Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative ^{31}P nuclear magnetic resonance method

J.R. Belmonte-Sánchez, L.M. Aguilera-Sáez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich. *Microchem. J.* 150 (2019).

Title: Determination of etidronic acid in vegetable-washing water by a simple and validated quantitative ^{31}P nuclear magnetic resonance method

Author: José Raúl Belmonte-Sánchez, Luis Manuel Aguilera-Sáez, Roberto Romero-González, José Luis Martínez Vidal, Francisco Javier Arrebola, Antonia Garrido Frenich

Publication: Microchemical Journal

Publisher: Elsevier

Date: November 2019

© 2019 Elsevier B.V. All rights reserved.

Logged in as:
José Raúl Belmonte
LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK **CLOSE WINDOW**

Copyright © 2019 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customer@copyright.com





7 CAPÍTULO SÉPTIMO. OTRAS APORTACIONES CIENTÍFICAS

Durante el desarrollo de la presente Tesis Doctoral, he llevado a cabo las siguientes aportaciones científicas, no incluidas en la misma:

Actividades de I+D+i

- MSCA-NIGHT- European Researchers' Night. Comisión Europea en la convocatoria Marie Sklodowska-Curie Actions. FUNDACION ANDALUZA PARA LA DIVULGACION DE LA INNOVACION Y EL CONOCIMIENTO. Participante colaborador de la actividad "La Química de los Alimentos" organizada para la Noche Europea .de los Investigadores 2018. Almería, Andalucía, España. 17/10/2018.
- MSCA-NIGHT- European Researchers' Night. Comisión Europea en la convocatoria Marie Sklodowska-Curie Actions. FUNDACION ANDALUZA PARA LA DIVULGACION DE LA INNOVACION Y EL CONOCIMIENTO. Participante colaborador de la actividad "La Química de los Alimentos" organizada para la Noche Europea .de los Investigadores 2017. Almería, Andalucía, España. 17/10/2017.
- MSCA-NIGHT- European Researchers' Night. Comisión Europea en la convocatoria Marie Sklodowska-Curie Actions.

FUNDACION ANDALUZA PARA LA DIVULGACION DE LA INNOVACION Y EL CONOCIMIENTO. Participante colaborador de la actividad "La Química de los Alimentos" organizada para la Noche Europea .de los Investigadores 2016. Almería, Andalucía, España. 17/10/2016.

- Reunión de la Sociedad Andaluza de Química Analítica (GRASEQA). Almería, Andalucía, España. 30/06/2016-01/07/2016. Participación en el Comité Organizador

Congresos nacionales e internacionales

- ¹H-NMR Y FUSIÓN DE DATOS COMO HERRAMIENTA METABOLÓMICA PARA LA CLASIFICACIÓN DE RONOS DORADOS MEDIANTE ANÁLISIS ESTADÍSTICOS MULTIVARIANTES. Belmonte Sánchez, José Raúl; Arrebola, Javier; Romero-González, Roberto; Martínez Vidal, José Luis; Garrido Frenich, Antonia. 2º CONGRESO de JÓVENES INVESTIGADORES en CIENCIAS AGROALIMENTARIAS UNIVERSIDAD DE ALMERÍA. Universidad de Almería. Almería. España. 17/10/19.
- CLASIFICACIÓN DE RONOS MEDIANTE MICROEXTRACIÓN EN FASE SÓLIDA DE ESPACIO EN CABEZA ACOLADA A GC-MS. Belmonte Sánchez, José Raúl; Arrebola, Javier; Romero-González, Roberto; Martínez Vidal, José Luis; Garrido Frenich, Antonia. 1er CONGRESO de



JÓVENES INVESTIGADORES en CIENCIAS
AGROALIMENTARIAS UNIVERSIDAD DE ALMERÍA.
Universidad de Almería. Almería. España. 20/12/18.

- RESONANCIA MAGNÉTICA NUCLEAR CUANTITATIVA DE FÓSFORO COMO HERRAMIENTA ANALÍTICA PARA LA DETERMINACIÓN DE ÁCIDO ETIDRÓNICO EN AGUAS DE LAVADO DE ALIMENTOS. J. R. Belmonte-Sánchez, R. Romero-González, F. J. Arrebola-Liébanas, J.L. Martínez-Vidal, A. Garrido Frenich. XVI REUNION DEL GRUPO REGIONAL ANDALUZ DE LA SOCIEDAD ESPAÑOLA DE QUÍMICAANALÍTICA (GRASEQA 2018). Palacio de los Congresos de Granada. Granada. España. 04/10/18-05/10/18.
- APLICACIÓN DE HERRAMIENTAS METABOLÓMICAS PARA LA CLASIFICACIÓN DE RONES DORADOS MEDIANTE CROMATOGRAFÍA DE LÍQUIDOS DE ULTRA ALTA PRESIÓNACOPLADA A ESPECTROMETRÍA DE MASAS ORBITRAP. J. R. Belmonte-Sánchez, R. Romero-González, F. J. Arrebola-Liébanas, J.L. Martínez-Vidal, A. Garrido Frenich. XVI REUNION DEL GRUPO REGIONAL ANDALUZ DE LA SOCIEDAD ESPAÑOLA DE QUÍMICAANALÍTICA (GRASEQA 2018). Palacio de los Congresos de Granada. Granada. España. 04/10/18-05/10/18.

- APLICACIÓN DE HERRAMIENTAS METABOLÓMICAS PARA LA CLASIFICACIÓN DE RONES DORADOS MEDIANTE EL USO DE UHPLC-ORBITRAP-MS. J. R. Belmonte-Sánchez, R. Romero-González, F. J. Arrebola-Liébanas, J.L. Martínez-Vidal, A. Garrido Frenich. VII Simposio de Investigación en Ciencias Experimentales. Universidad de Almería. Almería. España. 14/11/18-15/11/18.
- RHUM CLASSIFICATION USING FINGERPRINTING ANALYSIS BY HEAD SPACE SOLID PHASE MICROEXTRACTION COUPLED TO GAS CHROMATOGRAPHY-MASS SPECTROMETRY. José Raúl Belmonte-Sánchez, Javier Arrebola, Roberto Romero-González, Antonia Garrido Frenich, José Luis Martínez Vidal. Universidad de Valencia, Valencia, España. 05/07/18-07/07/18.
- BORON- ^{11}NMR AS A METHOD FOR QUANTITATIVE BORON ANALYSIS AND ITS APPLICATION TO DETERMINE BORIC ACID IN COMMERCIAL BIOCIDES. Belmonte-Sánchez J.R., Arrebola F.J., Romero-González R., Martínez-Vidal J.L., Garrido-Frenich A., Aguilera-Sáez L.M., Fernández I. Institute for Food Analysis and Research (IIAA), Universidad de Santiago de Compostela (USC). Santiago de Compostela. España. 16/06/18-22/06/18.



- AN INNOVATIVE METABOLOMIC APPROACH FOR GOLDEN RUMS CLASSIFICATION COMBINING ULTRA-HIGH PERFORMANCE LIQUID CHROMATOGRAPHY COUPLED TO ORBITRAP MASS SPECTROMETRY WITH CHEMOMETRIC METHODS. J. R. Belmonte-Sánchez, R. Romero-González, F. J. Arrebola-Liébanas, J.L. MartínezVidal, A. Garrido-Frenich. 14th ANNUAL LC/MS/MS WORKSHOP ON ENVIRONMENTAL APPLICATIONS AND FOOD SAFETY. Instituto de Estudios Catalanes, Barcelona. España. 26/06/18.
- FAST DETERMINATION OF MYCOTOXINS IN EDIBLE VEGETABLE OILS BY ULTRA HIGHPERFORMANCE LIQUID CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY. José L. Hidalgo-Ruiz, J.R. Belmonte-Sánchez, David Muñoz Jiménez, Roberto Romero-González, José Luis Martínez Vidal, Antonia Garrido Frenich, Jesús Alonso Barado, María Tello Liébana. 14th ANNUAL LC/MS/MS WORKSHOP ON ENVIRONMENTAL APPLICATIONS AND FOOD SAFETY. Instituto de Estudios Catalanes, Barcelona. España. 26/06/18.
- SOLID PHASE MECROEXTRACTION FOR FULLY AUTOMATED DETERMINATION OF ORGANIC POLLUTANTS BY GAS CHROMATOGRAPHY COUPLED

TO HIGH RESOLUTION MASS SPECTROMETRY IN ENVIRONMENTAL WATERS. I. Dominguez, F. J. Arrebola, M. Lopez-García, J. R. Belmonte-Sánchez, A. Garrido Frenich,, J. L. Martínez Vidal 19th International Symposium on Advances in Extraction Technologies. ExTech2017. Universidad de Santiago de Compostela (USC). Santiago de Compostela. España. 27/06/17-30/07/17.

- RUM CLASSIFICATION USING FINGERPRINTING ANALYSIS OF VOLATILE FRACTION BY HEADSPACE SOLID PHASE MICROEXTRACTION COUPLED TO GAS CHROMATOGRAPHY-MASS SPECTROMETRY VI MINISIMPOSIO DE INVESTIGACIÓN EN CIENCIAS EXPERIMENTALES. J. R. Belmonte Sánchez, F. J. Arrebola Liébanas, R. Romero González, J. L. Martínez Vidal, A. Garrido Frenich Universidad de Almería, Almería, España. 15/11/17.
- APPLICATION OF HEAD SPACE SOLID PHASE MICROEXTRACTION COUPLED TO GAS CHROMATOGRAPHY-MASS SPECTROMETRY TO THE STUDY OF RUMS. J. R. Belmonte-Sánchez, A. Garrido Frenich, F. J. Arrebola, J. L. Martínez Vidal, R. Romero González. 19th International Symposium on Advances in Extraction Technologies. ExTech2017. Universidad de Santiago de Compostela (USC). Santiago de Compostela. España. 27/0/17-30/07/17.



- RUM CLASSIFICATION USING FINGERPRINTING ANALYSIS OF VOLATILE FRACTION BY HEADSPACE SOLID PHASE MICROEXTRACTION COUPLED TO GAS CHROMATOGRAPHY-MASS SPECTROMETRY. J. R. Belmonte Sánchez, F. J. Arrebola Liébanas, R. Romero González, J. L. Martínez Vidal, A. Garrido Frenich. VI MINISIMPOSIO DE INVESTIGACIÓN EN CIENCIAS EXPERIMENTALES. Universidad de Almería, Almería. España. 15/11/17.

