

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

Departamento de Química Analítica

Centro Tecnológico de Investigación y Desarrollo del Alimento Funcional



**DESARROLLO DE NUEVAS ESTRATEGIAS PARA LA CARACTERIZACIÓN
E IDENTIFICACIÓN DE COMPUESTOS BIOACTIVOS EN FUENTES VEGETALES**

presentada por:

Aránzazu Morales Soto

para optar al grado de:

Doctor Internacional por la Universidad de Granada

Tesis doctoral dirigida por:

D. Alberto Fernández Gutiérrez

D. Antonio Segura Carretero

Dña. Cristina María Roldán Segura

Granada, 2014

Editor: Editorial de la Universidad de Granada
Autor: Aránzazu Morales Soto
D.L.: GR 2019-2014
ISBN: 978-84-9083-210-3

Esta tesis doctoral ha sido realizada gracias a una beca predoctoral concedida por la Junta de Andalucía (P09-CTS-4564) y a la financiación con cargo a fondos del Centro Tecnológico de Investigación y Desarrollo del Alimento Funcional (CIDAF) procedentes de diferentes proyectos, contratos y subvenciones de las Administraciones central y autonómica, plan propio de investigación de la UGR, así como de empresas interesadas en los resultados de la investigación.




**DESARROLLO DE NUEVAS ESTRATEGIAS PARA LA CARACTERIZACIÓN
E IDENTIFICACIÓN DE COMPUESTOS BIOACTIVOS EN FUENTES VEGETALES**

Memoria para optar al grado de Doctor Internacional por la UGR

POR

Aránzazu Morales Soto

Visado en Granada a 20 de Mayo de 2014



Fdo.: Prof. Dr. D. Alberto Fernández Gutiérrez
Catedrático del Departamento de Química Analítica
Facultad de Ciencias. Universidad de Granada



Fdo.: Prof. Dr. D. Antonio Segura Carretero
Catedrático del Departamento de Química Analítica
Facultad de Ciencias. Universidad de Granada



Fdo.: Dra. Dña. Cristina María Roldán Segura
Doctora en Ciencias Físicas por la Universidad de Granada
CIDAF, Centro Tecnológico de Investigación y Desarrollo del Alimento Funcional



Fdo.: Aránzazu Morales Soto



El Prof. Dr. D. ALBERTO FERNÁNDEZ GUTIÉRREZ, Catedrático de Universidad en el Departamento de Química Analítica "Profesor Fermín Capitán" de la Universidad de Granada y Director del Centro Tecnológico de Investigación y Desarrollo del Alimento Funcional (CIDAF),

CERTIFICA

Que el trabajo que se presenta en esta tesis doctoral con el título **"Desarrollo de nuevas estrategias para la caracterización e identificación de compuestos bioactivos en fuentes vegetales"** se ha realizado bajo mi dirección, la del Catedrático D. Antonio Segura Carretero y la Doctora Dña. Cristina María Roldán Segura, en los laboratorios del Centro Tecnológico de Investigación y Desarrollo del Alimento Funcional (Parque Tecnológico de Ciencias de la Salud), así como también de manera parcial en las instalaciones del Departamento de Nutrición y Ciencia de los Alimentos de la Universidad de Reading (Reino Unido), y que reúne todos los requisitos legales, académicos y científicos para hacer que el doctorando Dña. Aránzazu Morales Soto pueda optar al grado de Doctor Internacional por la Universidad de Granada.

Y para que así conste, expido y firmo el presente certificado en Granada a 20 de Mayo de 2014:



El doctorando Aránzazu Morales Soto y los directores de la tesis D. Alberto Fernández Gutiérrez, D. Antonio Segura Carretero y Dña. Cristina María Roldán Segura garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 20 de Mayo de 2014,

Director/es de la Tesis



Fdo.: D. Alberto Fernández
Gutiérrez



Fdo.: D. Antonio Segura
Carretero



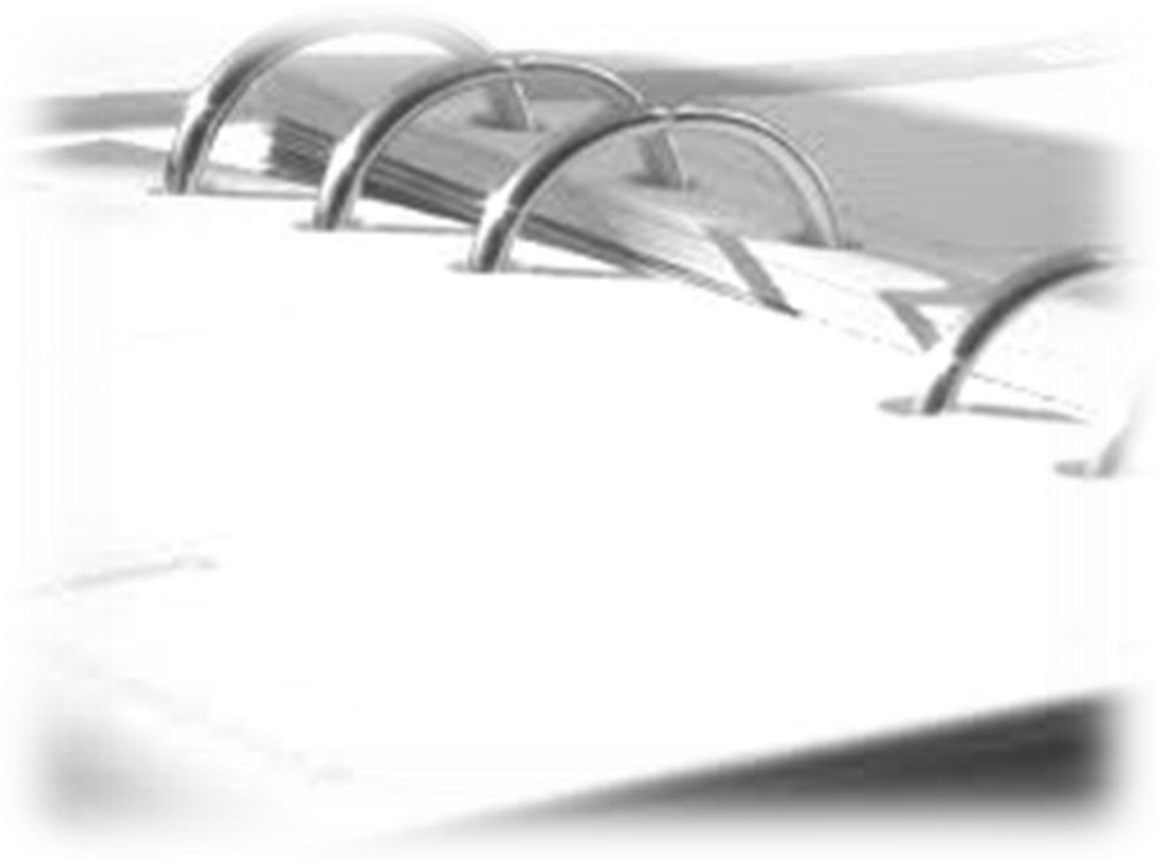
Fdo.: D. Cristina María
Roldán Segura

Doctoranda



Fdo.: Aránzazu Morales Soto





ÍNDICE GENERAL

RESUMEN.....	19
SUMMARY.....	29
OBJETIVOS	37
OBJECTIVES	43
INTRODUCCIÓN	49
1. Búsqueda de compuestos bioactivos a partir de fuentes vegetales.	51
2. Compuestos bioactivos de origen vegetal.	52
2.1. Terpenos.	52
2.2. Compuestos fenólicos.....	55
3. Bioactividad de terpenos y compuestos fenólicos.	58
3.1. Actividad antioxidante.	59
3.2. Actividad antimicrobiana.	65
3.3. Actividad anticancerígena.....	68
4. Extracción de compuestos bioactivos en muestras vegetales.....	71
5. Separación de compuestos bioactivos en muestras vegetales.	78
5.1. Cromatografía de gases (GC).	79
5.2. Cromatografía líquida (LC).	81
6. Detección de compuestos bioactivos en muestras vegetales: espectrometría de masas (MS).	84
7. Procesamiento convencional de los datos de espectrometría de masas (MS). ...	94
8. Nuevas herramientas de procesamiento de datos de espectrometría de masas (MS).	97
9. Identificación de compuestos bioactivos.....	103



METODOLOGÍA	107
A) Selección de muestras vegetales.	111
A.1) Selección de frutas y vegetales.	111
A.2) Selección de especies del género <i>Cistus</i>	114
A.3) Selección de muestras de aceite de oliva virgen extra.	115
B) Extracción de compuestos bioactivos.	118
B.1) Extracción de compuestos antioxidantes de frutas y vegetales.	118
B.2) Extracción de compuestos volátiles y compuestos fenólicos de <i>Cistus</i>	119
B.3) Extracción de compuestos fenólicos de aceite de oliva.	123
C) Análisis de composición química.	124
C.1) Análisis y caracterización de compuestos fenólicos de pimiento.	124
C.2) Análisis y caracterización compuestos volátiles y compuestos fenólicos de <i>Cistus</i>	127
C.3) Análisis de compuestos fenólicos de aceite de oliva.	132
D) Análisis de bioactividad.	134
D.1) Evaluación de capacidad antioxidante.	134
D.2) Evaluación de capacidad antimicrobiana.	137
D.3) Evaluación de capacidad anticancerígena.	139
E) Tratamiento de los datos de cromatografía acoplada a espectrometría de masas (LC-MS).	140
E.1) Fundamentos LC-MS desde un punto de vista de teoría de la señal.	142
E.2) Nuevas herramientas de tratamiento de datos LC-MS: métodos de <i>clustering</i>	172



F) Relación entre datos de espectrometría de masas y bioactividad.....	220
RESULTADOS Y DISCUSIÓN.....	243
Bloque 1. Caracterización e identificación de compuestos antioxidantes en muestras de frutas y vegetales.....	247
1.1. Evaluación de la capacidad antioxidante de cuarenta y cuatro variedades de frutas y vegetales cultivados en el sur de España.....	248
1.2. Caracterización de polifenoles y otros compuestos polares en tres muestras de pimiento (<i>Capsicum annun L.</i>).....	271
1.3. Aplicación de la herramienta <i>clustering</i> de líneas y la herramienta de correlación a las muestras de pimiento para la caracterización y la identificación de compuestos bioactivos.	287
Bloque 2. Caracterización e identificación de compuestos antimicrobianos en muestras de <i>Cistus</i>	311
2.1. Estudio de la fracción apolar de especies del género <i>Cistus</i>	313
2.2. Estudio de la fracción polar de especies del género <i>Cistus</i>	332
2.2.1. Capacidad antimicrobiana de los extractos polifenólicos de <i>Cistus</i> . .	333
2.2.2. Caracterización del perfil polifenólico de los extractos de <i>Cistus</i>	339
2.2.3. Aplicación de la herramienta <i>clustering</i> de líneas a la caracterización pormenorizada de los extractos de <i>Cistus</i> S2, S2a y S4.	345
2.2.4. Aplicación de la herramienta de correlación a la identificación de compuestos antimicrobianos de extractos polifenólicos de <i>Cistus</i>	364
Bloque 3. Caracterización e identificación de compuestos anticancerígenos en muestras de aceite de oliva.....	377
Characterization and identification of potentially anticarcinogenic compounds in olive oil samples.	379



3.1. Characterization of phenolic compounds in olive oil samples through the <i>line-clustering</i> method.	381
3.2. Identification of bioactive compounds through the correlation method between the MS and the bioactivity data.	387
3.2.1. Correlation with the rectangular <i>clustering</i>	390
3.2.2. Correlation with the <i>line-clustering</i>	394
3.2.3. Model for prediction of the bioactivity.	395
CONCLUSIONES.....	401
CONCLUSIONS.....	413
ABREVIATURAS	423
ÍNDICE FIGURAS Y TABLAS.....	429
APÉNDICE.....	445





RESUMEN

En esta memoria se reúnen los resultados obtenidos durante la realización de la tesis doctoral titulada "**Desarrollo de nuevas estrategias para la caracterización e identificación de compuestos bioactivos en fuentes vegetales**". En ella, se analiza una gran variedad de especies vegetales como frutas, hortalizas y otras plantas, como potenciales fuentes de compuestos bioactivos. También se estudia la información obtenida con espectrometría de masas mediante el desarrollo de nuevas herramientas de tratamiento de datos y su aplicación en la caracterización e identificación de compuestos bioactivos. La memoria se presenta dividida en tres secciones: introducción, metodología y resultados con su discusión.

En la **introducción**, se resumen las etapas del procedimiento analítico, haciendo hincapié en los compuestos, propiedades bioactivas y técnicas analíticas de mayor relevancia para la presente tesis. En primer lugar, se dedica un apartado a la búsqueda de posibles fuentes de compuestos bioactivos, que puedan ser interesantes en cuanto a su riqueza en composición y bioactividad. A continuación, se presentan los compuestos bioactivos divididos en dos familias, terpenos y compuestos fenólicos. Se incluye un apartado de bioactividad en el que se detallan las actividades biológicas antioxidante, antimicrobiana y anticancerígena. En cada caso, se describe el fundamento en que se basan los ensayos de valoración de la bioactividad. Posteriormente, se describen los métodos de extracción de compuestos bioactivos que pueden llevarse a cabo en función de la naturaleza de la muestra y el tipo de compuesto. Siguiendo el orden del procedimiento analítico, se describe el fundamento de las técnicas separativas utilizadas, cromatografía de gases y cromatografía líquida, así como el fundamento del sistema de detección empleado, la espectrometría de masas. Finalmente, se describe la sistemática seguida convencionalmente para procesar los datos de espectrometría de masas, exponiendo las limitaciones de



la misma y la consiguiente necesidad de desarrollar nuevas herramientas más versátiles para el tratamiento de dichos datos.

En la sección de **metodología**, se describe el procedimiento metodológico seguido durante las tesis. Este procedimiento general se divide en tantos apartados como etapas se llevan a cabo: (A) selección de fuentes vegetales, (B) extracción de los compuestos de interés, (C) análisis de la composición de las muestras seleccionadas, (D) evaluación de la bioactividad de las mismas, (E) tratamiento de los datos analíticos obtenidos y, por último, (F) correlación entre los datos analíticos tratados y los datos de bioactividad.

Así, se relata la diversidad de muestras estudiadas, de métodos de extracción utilizados, de técnicas analíticas y ensayos de bioactividad empleados, así como los reactivos y otros materiales utilizados durante el desarrollo experimental de la tesis. Especial mención se da a los apartados E y F, que incluyen una descripción detallada de las herramientas de tratamiento de datos de espectrometría de masas desarrolladas en la presente tesis. El tratamiento propuesto de los datos se denomina *clustering* y va precedido de una etapa de preprocesamiento que incluye eliminación de ruido. La aplicación del *clustering* permite la obtención automática de una lista de posibles compuestos, de gran ayuda para la caracterización. La estrategia de identificación de compuestos bioactivos desarrollada en esta tesis se basa en la correlación entre los datos de espectrometría de masas y los datos de bioactividad de un amplio conjunto de muestras.

Cada uno de los apartados de la metodología (A - F) se divide a su vez en tres subapartados, donde se recogen las diferentes especificaciones derivadas de la aplicación de la metodología general a cada estudio realizado, es decir, un primer estudio sobre vegetales y capacidad antioxidante, un segundo estudio



sobre *Cistus* y capacidad antimicrobiana y, por último, un tercer estudio sobre aceite de oliva y actividad anticancerígena.

La sección de **resultados y discusión** aparece dividida en tres bloques, donde se presentan los resultados de los estudios realizados.

En el bloque 1, se valoró en primer lugar la capacidad antioxidante de una amplia colección de frutas y vegetales del sur de España (principalmente, Andalucía), incluyendo diferentes variedades de las siguientes especies vegetales: aguacate, caqui, chirimoya, granada, limón, mango, melón, membrillo, níspero, sandía, uva, ajo, alcachofa, berenjena, calabacín, cebolla, espárrago, haba, judía verde, lechuga, patata, pimiento, tomate y zanahoria. Para valorar la capacidad antioxidante, se utilizaron tres métodos diferentes: el TEAC (trolox equivalent antioxidant capacity), el FRAP (ferric ion reducing antioxidant power) y el ORAC (oxygen radical absorbance capacity). Este trabajo ofrece un estudio comparativo de la capacidad antioxidante de estas variedades a lo largo del año, dado que fueron recolectadas a diferentes tiempos. Se ha de destacar principalmente el gran número de muestras implicado en el estudio. En segundo lugar, y en base a los resultados de capacidad antioxidante, se seleccionaron tres variedades de pimiento (*Capsicum annuum* L.) de alta capacidad antioxidante para un estudio de caracterización. Los extractos polifenólicos de pimiento se analizaron mediante cromatografía líquida de alta resolución acoplada a un sistema de detección de batería de diodos y a espectrometría de masas con analizador de tiempo de vuelo, mediante interfase electrospray (HPLC-DAD-ESI-TOF-MS). El procesamiento convencional de los datos, basado en los perfiles cromatográficos, permitió la identificación de 45 compuestos pertenecientes a diferentes tipos de metabolitos, de los cuales, 34 se describen por primera vez en esta matriz alimentaria. Los resultados sitúan al fruto del pimiento como una fuente potencial de compuestos fenólicos. Por



último en este primer bloque, se evaluaron las nuevas herramientas de tratamiento y procesamiento de datos de espectrometría de masas mediante la aplicación de las mismas con los datos analíticos y de bioactividad de las muestras de pimiento. La caracterización realizada a partir de los datos tratados con el *clustering* demostró que la herramienta es capaz de detectar de una manera rápida y automática una gran cantidad de compuestos entre los que se encuentran los 45 compuestos identificados mediante caracterización convencional. Se aplicó una variante de la herramienta de correlación a este estudio de pocas muestras, con la que se obtuvieron candidatos a ser responsables de la capacidad antioxidante.

En el bloque 2, se estudió tanto la fracción volátil como la fracción fenólica de especies españolas del género *Cistus*. Para la caracterización de los compuestos volátiles, se analizaron, mediante cromatografía de gases acoplada a espectrometría de masas (GC-MS), los compuestos volátiles previamente extraídos por microextracción en fase sólida del espacio en cabeza de las muestras (HS-SPME). Se identificaron 112 compuestos volátiles, la gran mayoría de naturaleza terpénica y de los cuales, 28 son descritos por primera vez en esta planta. Este estudio se llevó a cabo durante la estancia en el Departamento de Ciencias de la Nutrición y la Alimentación (Department of Food & Nutritional Sciences) de la Universidad de Reading, Reino Unido. El estudio de la fracción fenólica incluyó tanto la caracterización como la valoración de la actividad antimicrobiana de extractos derivados de las especies de *Cistus*. Los extractos obtenidos por diferentes métodos de extracción y secado se analizaron mediante cromatografía líquida de alta resolución acoplada a espectrometría de masas con analizador de tiempo de vuelo mediante sistema de ionización por electrospray (HPLC-ESI-TOF-MS). Se realizó la caracterización por el método convencional, lográndose identificar un total de 58 compuestos, de los cuales 18 se describen por primera vez en *Cistus*. Para evaluar la capacidad



antimicrobiana de los extractos polifenólicos de *Cistus*, se utilizaron dos modelos de bacteria, *Escherichia coli* como modelo de bacteria Gram negativa y *Staphylococcus aureus* como bacteria Gram positiva. Los resultados mostraron que la actividad biológica no sólo dependió del quimiotipo de la especie utilizada, sino que también influyeron los procedimientos de extracción y secado utilizados. El estudio de la capacidad antimicrobiana se realizó en colaboración con el Instituto de Biología Molecular y Celular de la Universidad Miguel Hernández de Elche (Alicante, España). Por último en este segundo bloque, se evaluaron las herramientas de *clustering* y correlación mediante su aplicación con los datos analíticos y de bioactividad de los extractos polifenólicos de *Cistus*. Se seleccionaron los tres extractos de *Cistus* de mayor bioactividad y se analizaron mediante cromatografía líquida de ultra alta resolución acoplada a espectrometría de masas con analizador de masas de cuadrupolo-tiempo de vuelo mediante una interfase electrospray (UPLC-ESI-QTOF-MS). Los datos procedentes de este análisis se trataron mediante el *clustering*, poniendo de manifiesto, por un lado, la información estructural proporcionada por el QTOF y, por otro, la utilidad del *clustering* en la caracterización pormenorizada de extractos. El tratamiento de los datos permitió la detección de picos que por coelución, se habían obviado en la caracterización convencional anterior. Para el estudio de correlación, se utilizó una colección de 26 extractos polifenólicos de *Cistus* y sus correspondientes datos de capacidad antimicrobiana, correlacionando de manera separada los resultados realizados con los diferentes modelos de bacteria. En ambos casos, se obtuvieron candidatos a ser responsables de la bioactividad, destacando principalmente los elagitaninos frente a bacterias Gram negativas y los flavonoides frente a bacterias Gram positivas.

En el bloque 3, se utilizaron los datos analíticos procedentes de una amplia colección de extractos polifenólicos de aceite de oliva virgen extra.



Primeramente, se realizó la caracterización de los extractos a partir de los datos tratados con el *clustering*. De esta forma, se demostró que se obtienen los compuestos identificados en estudios de caracterización anteriores, realizados con el método convencional. En segundo lugar, se realizó el estudio de correlación con la bioactividad. Para ello, se valoró la actividad anticancerígena de los extractos polifenólicos de aceite frente a la línea celular de cáncer de mama JIMT-1, en colaboración con el Instituto Catalán de Oncología de Girona. La correlación con la bioactividad se realizó partiendo de los datos tratados mediante dos tipos de *clustering*. En ambos casos, los candidatos obtenidos como compuestos responsables de la actividad anticancerígena observada fueron tres isómeros de la oleuropeína aglicona. La similitud en los resultados obtenidos con ambos métodos de *clustering* pone de manifiesto la utilidad de la herramienta de correlación, a la vez que se demuestra la rapidez y robustez del *clustering*. Todo ello permitió verificar la potencialidad de la herramienta de correlación propuesta como estrategia para la identificación de compuestos candidatos a ser responsables de la bioactividad, así como también la potencialidad predictiva de la misma.





SUMMARY

This work summarizes all the results presented in the PhD thesis entitled "**Desarrollo de nuevas estrategias para la caracterización e identificación de compuestos bioactivos en fuentes vegetales (*Development of new strategies for characterization and identification of bioactive compounds in plant sources*)**". In this thesis, a wide variety of vegetal species such as fruits, vegetables and other plants are assessed as potential sources of bioactive compounds. The potential offered by the mass spectrometry information is also studied in this work, through the development of new data processing tools and their application to characterization and identification of bioactive compounds. The current report is divided into three main sections: Introduction, Methodology, and Results and Discussion.

The **Introduction** deals with the stages of the analytical procedure of a sample, highlighting the most relevant compounds, bioactive properties and analytical techniques used in this thesis. A first subsection about the search for potential sources of bioactive compounds is included. Afterwards, terpenes and phenolic compounds are described due to their role as bioactive compounds. A subsection about the biological activity of these compounds is described next, focused on antioxidant, anticancer, and antimicrobial activities. For every case, the fundamentals of the bioactivity assays are described. Subsequently, the different extraction systems that can be used for bioactive compounds are itemized depending on the compounds and the nature of the sample. Continuing with the stages of the analytical procedure, gas chromatography and liquid chromatography separation techniques are described, as well as the mass spectrometry detection system used in this work. Finally, the method carried out to process the mass spectrometry data is described. Certain drawbacks of this conventional method imply the need to develop new and versatile methods for processing the mass spectrometry data.



The **Methodology** section describes the methodological procedure followed in the thesis. This general procedure is divided into as many subsections as steps are performed: (A) selection of plant sources, (B) extraction of the compounds of interest, (C) analysis of the composition of the selected samples, (D) bioactivity assessment thereof, (E) analytical data processing and, finally, (F) correlation between the processed analytical data and the bioactivity data.

This way, the wide variety of samples, extraction methods, bioactivity assays and analytical techniques are described, as well as the reagents and other materials employed in the present work. Special mention is given to the E and F subsections, including a detailed description of the new mass spectrometry data processing tools developed in this PhD thesis. The proposed tool for data treatment is called *clustering* and is preceded by a preprocessing stage including noise removal. The application of *clustering* automatically provides a list of possible compounds, which is very useful for characterization. The strategy for identifying bioactive compounds developed in this thesis is based on the correlation between the mass spectrometry data and bioactivity data of a large set of samples.

The subsections of the methodology (A - F) include the different specifications from the application of the general methodology to the studies performed: an initial study on vegetable and antioxidant capacity, a second study on *Cistus* and antimicrobial capacity and, finally, a third study on olive oil and anticancer activity.

The **Results and Discussion** section is composed by three subsections, corresponding to the studies carried out.

In Subsection 1, firstly the antioxidant capacity of a large collection of fruit and vegetable samples grown in southern Spain (mainly, Andalusia) was evaluated.



The samples included different cultivars of their species, collected at different times year round: asparagus, artichoke, carrot, eggplant, fava bean, garlic, green bean, lettuce, onion, pepper, potato, tomato, zucchini, avocado, custard apple, grape, lemon, loquat, mango, melon, persimmon, pomegranate, quince and watermelon. Three different methods were used for assessing the antioxidant capacity: TEAC (trolox equivalent antioxidant capacity), FRAP (ferric ion reducing antioxidant power) and ORAC (oxygen radical absorbance capacity). This work offers a comparative study on the antioxidant capacity provided by the intake of these Spanish fruits and vegetables year round. It must be highlighted the huge number of samples involved in the study. Secondly, three cultivars of pepper (*Capsicum annum* L.) showing high antioxidant capacity values were selected for a characterization study of their polar fraction. The polyphenolic pepper extracts were analyzed by using high performance liquid chromatography coupled to diode-array and time-of-flight mass spectrometry (HPLC-DAD-ESI-TOF-MS). The conventional data processing, based on chromatographic profiles, allowed the identification of 45 compounds belonging to different metabolite classes, 34 of which were described for the first time in pepper. The results showed that pepper is a good source of phenolic compounds and this information could be used in order to update Spanish food composition tables. Finally in this subsection, the new data processing tools were evaluated by applying them to the analytical and the bioactivity data of the pepper samples. The characterization through *clustering* enabled to detect in a quick and automatic manner a large number of compounds, among which the 45 compounds identified by the conventional characterization were found. For this study including few samples, the correlation tool was used with some variation and candidates to be responsible for the antioxidant capacity were obtained.

In Subsection 2, both the volatile and phenolic fractions of Spanish species of *Cistus* genus were studied. For the characterization of volatile fraction, gas



chromatography coupled to mass spectrometry (GC-MS) was used to analyze the volatile compounds extracted by using headspace solid-phase microextraction (HS-SPME). This analysis enabled the identification of 112 volatile compounds (mainly terpenes), 28 of which reported for the first time in *Cistus*. This study was carried out during the stay in the Department of Nutrition and Food Sciences in Reading University (United Kingdom). The study of the phenolic fraction included both the characterization and the evaluation of antimicrobial activity of the extracts from different *Cistus* species. Different extraction and drying systems were used to obtain the extracts, which were analyzed by using high performance liquid chromatography coupled to time-of-flight mass spectrometry (HPLC-ESI-TOF-MS). The conventional characterization of these extracts, a total of 58 compounds were tentatively identified, 18 of which were reported for the first time in *Cistus*. In order to assess the antimicrobial capacity of the polyphenolic *Cistus* extracts, two bacteria models were employed, i.e. *Escherichia coli* and *Staphylococcus aureus* as Gram negative and Gram positive bacteria models, respectively. The results showed that the biological activity not only depended on the chemotype of the species used, but also on the extraction and drying techniques used. This work was conducted in collaboration with the Institute of Molecular and Cellular Biology from Miguel Hernández University (Elche, Alicante, Spain). Finally in this subsection, the new data processing tools were evaluated by applying them to the analytical and the bioactivity data of the *Cistus* polyphenolic extracts. Three *Cistus* extracts showing the highest bioactivity were selected and analyzed by using ultra high performance liquid chromatography coupled to quadrupole-time-of-flight mass spectrometry (UPLC-ESI-QTOF-MS). The data from these analyses were processed by the *clustering* tool, demonstrating the structural information provided by QTOF mass analyzer, and the usefulness of the *clustering* for comprehensive characterization of extracts. The application of



clustering allowed the detection of peaks that were ignored in the conventional characterization due to co-elution. For the correlation study, the analytical data from a collection of 26 polyphenolic *Cistus* extracts and their antimicrobial capacity data were correlated, separately for the different bacteria models. In both cases, candidates for bioactive compounds were obtained, highlighting ellagitannins against Gram negative bacteria and flavonoids against Gram positive bacteria.

In Subection 3, the analytical data from a collection of polyphenolic extracts of extra-virgin olive oil were used. In the first place, a characterization of the extracts by using the *clustering* processing tool was performed. It was found that all compounds previously identified by the conventional procedure were obtained with this new tool. In the second place, the correlation study with the bioactivity data was implemented. For this purpose, the anticarcinogenic activity of the extracts against the JIMT-1 breast cancer cell line was evaluated *in vitro*. The study dealing with the anticancer activity was developed in collaboration with the Catalan Institute of Oncology (ICO) in Girona. The correlation study was performed by using the analytical data from two different *clustering* tools. In both cases, the candidates obtained as responsible compounds for the anticancer activity were three isomers of oleuropein aglycone. The similarity of the results obtained with both *clustering* tools showed the usefulness of the correlation tool, while the speed and robustness of the *clustering* was demonstrated. This confirmed the potential of the correlation tool as a strategy for identifying bioactive candidates as well as the predictive potential of this method.





OBJETIVOS

La búsqueda de fuentes naturales de origen vegetal, y de principios bioactivos a partir de éstas, es un campo de la ciencia muy activo en la actualidad, especialmente para la prevención y/o el tratamiento de enfermedades. En el estudio de un extracto vegetal, es importante conocer no sólo su actividad biológica, sino también los compuestos presentes en él. En cuanto a la caracterización de los compuestos del extracto, la metodología convencional basada en la identificación de picos cromatográficos posee ciertas limitaciones asociadas a la coelución de los picos. En lo referente a la bioactividad del extracto, es de gran interés conocer qué compuesto/s es/son el/los responsable/s de la misma. Para ello, convencionalmente se recurre al fraccionamiento del extracto, proceso que puede estar limitado por el tiempo requerido hasta obtener el compuesto bioactivo aislado y puro, así como por el coste de los ensayos de valoración de la bioactividad de cada fracción.

Por todo ello, la presente tesis doctoral pretende tres **objetivos generales**:

- Evaluación del potencial de frutas, vegetales y otras plantas como fuentes de compuestos bioactivos mediante estudios de caracterización de compuestos fenólicos y terpenos, así como estudios de valoración de diferentes propiedades biológicas (antioxidante, antimicrobiana y anticancerígena).
- Desarrollo de nuevas herramientas de tratamiento y procesamiento de datos de espectrometría de masas: por un lado, una herramienta de *clustering* basada en el preprocesado adecuado de los datos para la eliminación del ruido, el tratamiento de los datos en tres dimensiones (tiempo de retención, masa, intensidad) y la obtención de una lista que contenga todos los compuestos, y por otro lado, una herramienta de



correlación entre los datos de espectrometría de masas y los datos de bioactividad.

- Aplicación de las nuevas herramientas desarrolladas a diversos casos para su evaluación en estudios de caracterización pormenorizada de extractos e identificación de compuestos potencialmente responsables de la bioactividad.

Estos objetivos generales se concretan en nueve **objetivos específicos** que se abordarán durante el desarrollo experimental de la presente tesis:

- Determinación de la capacidad antioxidante de una colección de cuarenta y cuatro variedades andaluzas de frutas y vegetales mediante la utilización de tres métodos diferentes: TEAC (trolox equivalent antioxidant capacity), FRAP (ferric ion reducing antioxidant power) y ORAC (oxigen radical absorbance capacity).
- Caracterización del contenido en compuestos fenólicos y otros compuestos polares de distintas variedades de pimiento (*Capsicum annum* L.) mediante cromatografía líquida de alta resolución acoplada a diversos sistemas de detección, detector de batería de diodos (DAD) y espectrometría de masas con analizador de tiempo de vuelo e interfase electrospray (HPLC-DAD-ESI-TOF-MS).
- Aplicación de las herramientas de *clustering* y correlación a la caracterización de los extractos polifenólicos de pimiento y a la identificación de los compuestos potencialmente responsables de la capacidad antioxidante, respectivamente.
- Caracterización del perfil aromático de plantas del género *Cistus*, analizando, mediante cromatografía de gases acoplada a espectrometría



de masas (GC-MS), los compuestos volátiles del espacio de cabeza obtenidos mediante microextracción en fase sólida (HS-SPME).

- Caracterización mediante cromatografía líquida de alta resolución acoplada a espectrometría de masas con analizador de masas de tiempo de vuelo e interfase electrospray (HPLC-ESI-TOF-MS) y valoración de la capacidad antimicrobiana de extractos polifenólicos de diferentes especies españolas de *Cistus* mediante la utilización de modelos de bacteria Gram negativa y Gram positiva.
- Aplicación de la herramienta de *clustering* a la caracterización pormenorizada de tres de los extractos polifenólicos de *Cistus* más bioactivos, analizados mediante cromatografía líquida de ultra alta resolución acoplada a espectrometría de masas con analizador de masas cuadrupolo tiempo de vuelo (UPLC-ESI-QTOF-MS).
- Aplicación de la herramienta de correlación con los extractos polifenólicos de *Cistus* para la identificación de los compuestos potencialmente responsables de la actividad antimicrobiana, frente a bacterias Gram positivas y Gram negativas.
- Aplicación de la herramienta de *clustering* a la caracterización pormenorizada de extractos polifenólicos de aceite de oliva virgen extra.
- Aplicación de la herramienta de correlación a la identificación de los compuestos potencialmente responsables de la actividad anticancerígena frente a la línea celular de cáncer de mama JIMT-1, a partir de la colección de extractos de aceite de oliva.





OBJECTIVES

Currently, the search of natural plant sources of bioactive compounds is a very active field of Science, especially for the prevention and/or the treatment of diseases. In the study of a plant extract, it is important to know not only their biological activity, but also the compounds present. Regarding the characterization of the compounds of the extract, the conventional method based on the identification of chromatographic peaks has certain limitations associated with coelution. With regard to the bioactivity of the extract, it is interesting to determine which compound/s is/are responsible for the biological properties. Conventionally, this identification is based on fractionation, evaluation of the bioactivity associated with each fraction, and study of the compounds in the most bioactive one. The evaluation of the bioactivity allows the identification of the fraction containing the bioactive compound, but if more than one compound is found in that fraction, a new further fractionation is necessary in order to isolate and identify the bioactive compound. Available analytical techniques allow the development of advanced methods for the identification of bioactive compounds in complex samples. The effectiveness of these methods strongly depends on the complexity of the samples and on the effort involved on the fractionation and the evaluation of the bioactivity.

Therefore, the **main objectives** of this PhD thesis are:

- Evaluation of the potential of fruits, vegetables and other plants as sources of bioactive compounds through the characterization of phenolic and volatile compounds and the assessment of different biological activities (antioxidant, antimicrobial and anticarcinogenic).
- Development of new tools for processing mass spectrometry data: on the one hand, a *clustering* tool based on appropriate data pre-processing for noise removal, the three dimensions processing of the data (retention



time, mass, intensity) and obtaining a list of compounds; and on the other hand, a tool based on the correlation between the mass spectrometry data and the bioactivity data.

- Application of the new tools to several cases in order to evaluate the developed tools in both characterization and identification of potentially bioactive compounds studies.

These general aims resulted in these **specific objectives**:

- Determination of the antioxidant capacity of a collection of forty-four Andalusian cultivars of fruits and vegetables by using three different methods: TEAC (trolox equivalent antioxidant capacity), FRAP (ferric ion Reducing antioxidant power) and ORAC (oxygen radical absorbance capacity).
- Characterization of phenolic and other polar compounds from different cultivars of pepper (*Capsicum annuum* L.) by using high resolution liquid chromatography coupled to different detection systems, diode array detector (DAD) and mass spectrometry with time-of-flight analyzer and electrospray interface (HPLC-DAD-ESI-TOF-MS).
- Application of the *clustering* and correlation tools to the study of pepper extracts for the characterization of phenolic compounds and the identification of potentially antioxidant compounds, respectively.
- Characterization of the volatile profile of *Cistus* plants by using headspace solid phase microextraction and gas chromatography coupled to mass spectrometry (HS-SPME-GC-MS).



- Characterization of different polyphenolic extracts from Spanish *Cistus* species by using high resolution liquid chromatography coupled to mass spectrometry time-of-flight analyzer (HPLC-ESI-TOF-MS) and evaluation of their antimicrobial activity with Gram positive and Gram negative bacteria models.
- Application of the *clustering* tool to the comprehensive characterization of three of the most bioactive polyphenolic extracts from *Cistus*, analyzed by ultra high resolution liquid chromatography coupled to mass spectrometry quadrupole-time-of-flight mass analyzer (UPLC-ESI-QTOF-MS).
- Application of the correlation tool to a study involving a collection of twenty-six polyphenolic *Cistus* extracts in order to identify the compounds potentially responsible for the antimicrobial activity against Gram positive and Gram negative bacterium.
- Application of the *clustering* tool to the comprehensive characterization of polyphenolic extracts of extra virgin olive oil from different olive cultivars and Spanish locations.
- Application of the correlation tool to a study involving a collection of twenty-two olive oil polyphenolic extracts in order to identify the compounds potentially responsible for the anticancer activity against JIMT-1 breast cancer cell line.





INTRODUCCIÓN

1. BÚSQUEDA DE COMPUESTOS BIOACTIVOS A PARTIR DE FUENTES VEGETALES.

En el contexto del descubrimiento de compuestos con actividad biológica, el primer paso es la selección de posibles fuentes. La investigación en torno a los compuestos bioactivos se basa en la búsqueda de fuentes naturales que puedan aplicarse posteriormente no sólo en la industria farmacéutica y alimentaria, sino también en cosmética e higiene, implicando una disminución del impacto medioambiental. Otro punto crucial del objetivo de la búsqueda de compuestos bioactivos sería encontrar nuevas fuentes, más rentables y de mayor rendimiento. A este respecto, destaca el potencial de las fuentes de origen vegetal, teniendo en cuenta su biodiversidad y la amplia variedad de moléculas que producen¹⁻⁶.



Figura 1 Extracto vegetal

El proceso de identificación de compuestos bioactivos siempre está sujeto a estudios de relación estructura-bioactividad, por lo que el siguiente paso tras la selección de muestra es el estudio de la bioactividad y de la composición química. Las especies vegetales constituyen mezclas complejas de cientos de compuestos, entre los cuales uno o sólo unos pocos serán los responsables de la actividad biológica. De ahí, que las nuevas tendencias vayan encaminadas a la

¹ Gullo, V.P. *et al.*, 2006. Drug discovery from natural products. *Journal of Industrial Microbiology & Biotechnology*, 33(7), pp.523–531.

² Kingston, D.G.I., 2011. Modern natural products drug discovery and its relevance to biodiversity conservation. *Journal of Natural Products*, 74(3), pp.496–511.

³ Newman, D.J. & Cragg, G.M., 2007. Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products*, 70(3), pp.461–477.

⁴ Kalra, E.K., 2003. Nutraceutical-definition and introduction. *AAPS PharmSci*, 5(3), pp.27–28.

⁵ Raskin, I. *et al.*, 2002. Plants and human health in the twenty-first century. *Trends in Biotechnology*, 20(12), pp.522–531.

⁶ Peschel, W. *et al.*, 2006. An industrial approach in the search of natural antioxidants from vegetable and fruit wastes. *Food Chemistry*, 97(1), pp.137–150.

identificación del candidato más prometedor de la mezcla compleja, para su posterior aislamiento y evaluación⁷.

2. COMPUESTOS BIOACTIVOS DE ORIGEN VEGETAL.

Existe una gran variedad de compuestos bioactivos con distintas estructuras y actividades biológicas. Concretamente, en el reino vegetal se pueden distinguir cuatro grandes grupos de compuestos bioactivos, entre los cuales se incluyen las sustancias nitrogenadas, las azufradas, las terpénicas y las fenólicas. Estas dos últimas familias requieren una mención especial, por el hecho de ser objeto de estudio en la presente memoria. A continuación, se entra en detalle en cuanto a la clasificación y estructura de los compuestos terpénicos y los compuestos fenólicos⁸.

2.1. Terpenos.

Los terpenos constituyen uno de los grupos más importantes de compuestos volátiles de la plantas. Forman parte de los aceites esenciales de las mismas, como mezclas de varias sustancias químicas responsables del aroma característico de flores, árboles, frutos, hierbas, especias, etc⁹.

Los terpenos son metabolitos secundarios que se liberan como mecanismo de defensa de las plantas, tanto de sus hojas y flores, como de sus frutos. Estas funciones defensivas están relacionadas con las agresiones bióticas, como el ataque de herbívoros o las infecciones causadas por patógenos víricos, bacterianos o fúngicos. Se sintetizan a partir de metabolitos primarios por dos rutas: la del ácido mevalónico que tiene lugar en el citoplasma y en la que tres

⁷ Cordell, G. a. & Shin, Y.G., 1999. Finding the needle in the haystack. The dereplication of natural product extracts. *Pure and Applied Chemistry*, 71(6), pp.1089–1094.

⁸ Tomás-Barberán, F.A., 2003. Los polifenoles de los alimentos y la salud. *Alimentación, Nutrición y Salud*, 10, pp.41–53.

⁹ Burt, S., 2004. Essential oils: their antibacterial properties and potential applications in foods-a review. *International Journal of Food Microbiology*, 94(3), pp.223–253.



moléculas de acetil-CoA se condensan para formar ácido mevalónico que reacciona hasta formar isopentenil difosfato (IPP), o bien la ruta del metileritritol fosfato (MEP) que tiene lugar en los cloroplastos y genera también IPP^{10,11}.

Estructuralmente, los terpenos son hidrocarburos pertenecientes a las familias de los alquenos, alcoholes, ésteres, éteres, aldehídos y cetonas. Cuando las moléculas contienen átomos de oxígeno se da preferencia al término terpenoide, aunque ambas denominaciones, terpeno y terpenoide, suelen utilizarse indistintamente. También se emplea el término isoprenoide, por tratarse de derivados del precursor inmediato del isopreno. Por la unión sucesiva de dos, tres y cuatro unidades de isopreno, se generan, respectivamente, monoterpenos, sesquiterpenos y diterpenos, de volatilidad decreciente¹².

De esta forma, pueden clasificarse en función del número de unidades de isopreno (molécula de 5 átomos de carbono): los terpenos de 10 átomos de carbono contienen dos unidades de isopreno y se llaman monoterpenos; los de 15 átomos de carbono tienen tres unidades de isopreno y se denominan sesquiterpenos, y los de 20 átomos de carbono son los diterpenos, formados por cuatro unidades de isopreno. También existen triterpenos con 30 átomos de carbono, tetraterpenos con 40 átomos de carbono, y se habla de politerpenos cuando contienen más de 8 unidades de isopreno. En la Figura 2 se muestra la clasificación de los terpenos así como algunos ejemplos con su respectiva estructura.

¹⁰ Ávalos, A. & Elena, G., 2009. Metabolismo secundario de plantas. *Reduca Biología Serie Fisiología Vegetal*, 2(3), pp.119–145.

¹¹ Sharkey, T.D. & Yeh, S., 2001. Isoprene emission from plants. *Plant Molecular Biology*, 52, pp.407–436.

¹² Singer, A.C., Crowley, D.E. & Thompson, I.P., 2003. Secondary plant metabolites in phytoremediation and biotransformation. *Trends in Biotechnology*, 21(3), pp.123–130.



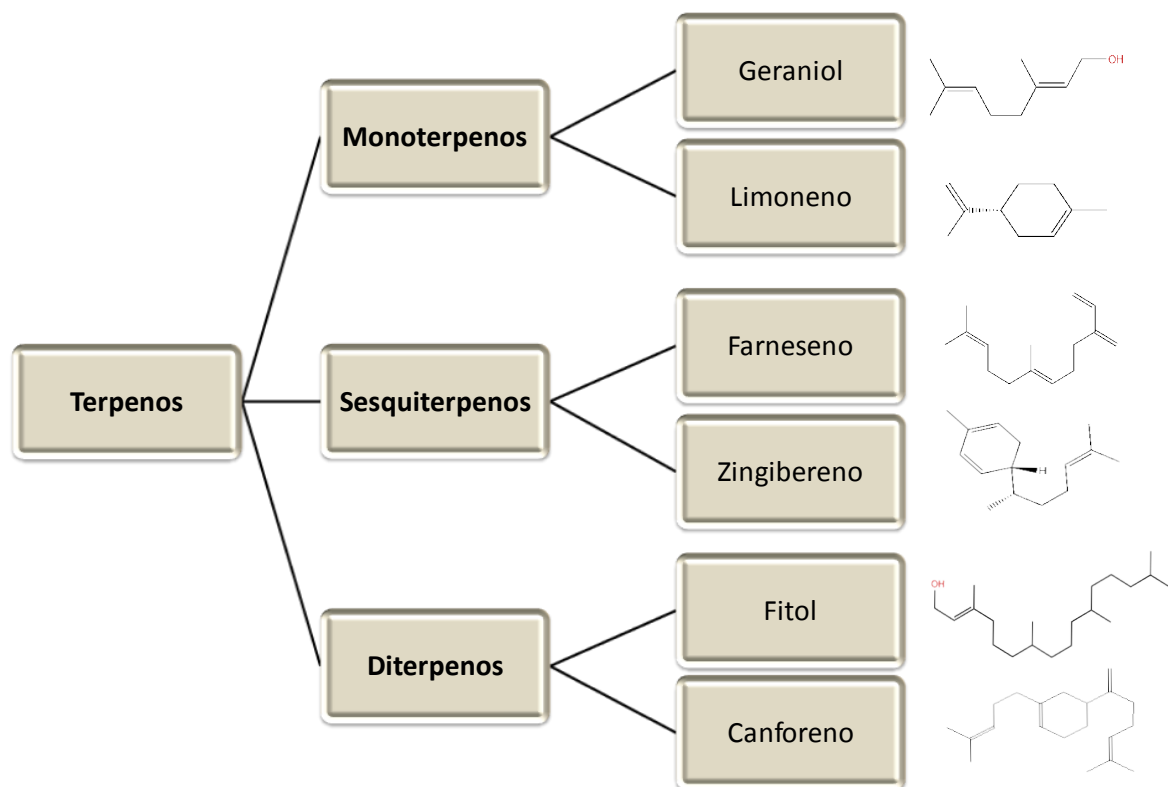


Figura 2. Clasificación y algunas estructuras de terpenos.

Los terpenos que se encuentran en los aceites esenciales de plantas son generalmente **monoterpenos**, como por ejemplo el limoneno y el mentol, principales constituyentes de los aceites de limón y menta, respectivamente. Los monoterpenos pueden ser acíclicos (lineales) o cíclicos. Entre los monoterpenos acíclicos, se incluyen los trienos mirceno y los alcoholes geraniol, nerol, linalol, etc. Dentro de los monoterpenos cíclicos, se pueden clasificar en monocíclicos como el limoneno, el terpineno, el felandreno y la pulegona y en bicíclicos, algunos ejemplos de los cuales son el pineno, el borneol y el alcanfor.

Como en el caso de los monoterpenos, los **sesquiterpenos** cuya fórmula molecular sería $C_{15}H_{24}$, también pueden ser acíclicos o cíclicos. Ejemplos de sesquiterpenos acíclicos pueden ser el farneseno y el farnesol. Entre los sesquiterpenos monocíclicos, destaca el zingibereno como un constituyente del

aceite esencial de jengibre. Los sesquiterpenos bicíclicos también son componentes importantes en los aceites esenciales de plantas, entre los que destacan el selineno del aceite de apio y el guayazuleno, previamente identificado en el aceite esencial de geranio. Este último compuesto es un ejemplo de la familia de los azulenos, que dentro de los sesquiterpenos bicíclicos reciben dicho nombre por ser los responsables del color azul de los aceites esenciales de ciertas plantas.

Los **diterpenos**, al poseer una estructura de mayor tamaño, requieren una mayor energía para volatilizarse. Por ello, es menos común encontrarlos en aceites esenciales. Sin embargo, suelen encontrarse en resinas como es el caso del ácido abiético. Como ejemplos, destacan el fitol, un alcohol diterpénico (acíclico) y el canforeno, un diterpeno cíclico¹³⁻¹⁵.

2.2. Compuestos fenólicos.

Los compuestos fenólicos presentan gran diversidad de estructuras químicas pero todas ellas contienen un anillo aromático con al menos un sustituyente hidroxilo. Constituyen uno de los grupos de sustancias más numerosos y ampliamente distribuidos en los vegetales, con más de 8.000 estructuras fenólicas conocidas hasta hoy¹⁶.

Al igual que los terpenos, los compuestos fenólicos son metabolitos secundarios de las plantas que desempeñan una función defensiva como respuesta al estrés ambiental como puede ser el ataque de patógenos, la contaminación del aire o

¹³ Croteau, R., Kutchan, T.M. & Lewis, N.G., 2000. Natural products (secondary metabolites). In B. Buchanan, W. Gruissem, & R. Jones, eds. *Biochemistry & Molecular Biology of Plants*. pp. 1250–1318.

¹⁴ Breitmaier, E., 2006. Terpenes: importance, general structure, and biosynthesis. *Terpenes: Flavors, Fragrances, Pharmaca, Pheromones*. pp. 1-9.

¹⁵ Humphrey, A.J., Beale, M.H., 2007. Terpenes. *Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet*. pp. 47-74.

¹⁶ Naczki, M. & Shahidi, F., 2006. Phenolics in cereals, fruits and vegetables: occurrence, extraction and analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 41(5), pp.1523–1542.



exposición a temperaturas extremas¹⁷. La biosíntesis de los polifenoles tiene lugar a través de dos rutas primarias: la ruta del ácido shikímico y la ruta del acetato. La ruta del ácido shikímico proporciona la síntesis de los aminoácidos aromáticos (fenilalanina o tirosina), y la síntesis de los ácidos cinámicos y sus derivados (fenoles sencillos, ácidos fenólicos, cumarinas, lignanos y derivados del fenilpropano). La ruta del acetato proporciona las quinonas y las xantonas¹⁸.

Existen varias clases y subclases de compuestos fenólicos, que se definen en función del número de anillos fenólicos que poseen y de los elementos estructurales que presentan estos anillos. Los grupos principales de compuestos fenólicos son: ácidos fenólicos (derivados del ácido hidroxibenzóico o del ácido hidroxicinámico), estilbenos, lignanos, alcoholes fenólicos y flavonoides (ver Figura 3)¹⁹.

¹⁷ Ellnain-Wojtaszek, M., Kruczynski, Z. & Kasprzak, J., 2001. Ellnain-Wojtaszek. *Acta Poloniae Pharmaceutica-Drug Research*, 58(3), pp.205–209.

¹⁸ Quiñones, M., Miguel, M. & Aleizandre, A., 2012. Los polifenoles , compuestos de origen natural con efectos saludables sobre el sistema cardiovascular. *Nutrición Hospitalaria*, 27(1), pp.76–89.

¹⁹ Bravo, L., 1998. Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. *Nutrition Reviews*, 56, pp. 317–333.



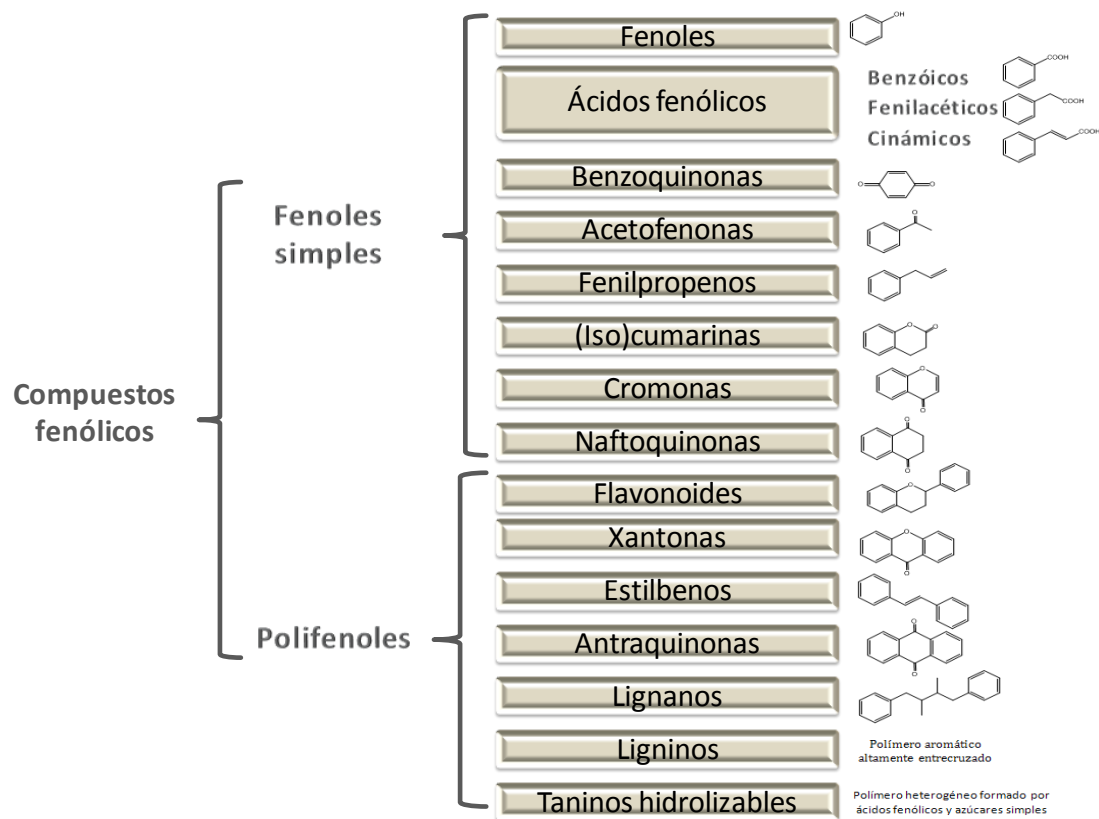


Figura 3. Clasificación y estructuras de los compuestos fenólicos.

De todas las familias de polifenoles, los flavonoides son los más ampliamente distribuidos en la naturaleza, llegando a alcanzar los dos tercios de los fenoles de la dieta. Son compuestos de bajo peso molecular que comparten un esqueleto común de difenilpirano ($C_6-C_3-C_6$), compuesto por dos anillos fenilo (A y B) ligados a través de un anillo C de pirano heterocíclico. Se pueden clasificar a su vez en trece subclases representadas en la Figura 4²⁰.

²⁰ Scalbert, A. & Williamson, G., 2000. Dietary intake and bioavailability of polyphenols. *The Journal of Nutrition*, 130, pp.2073–2085.

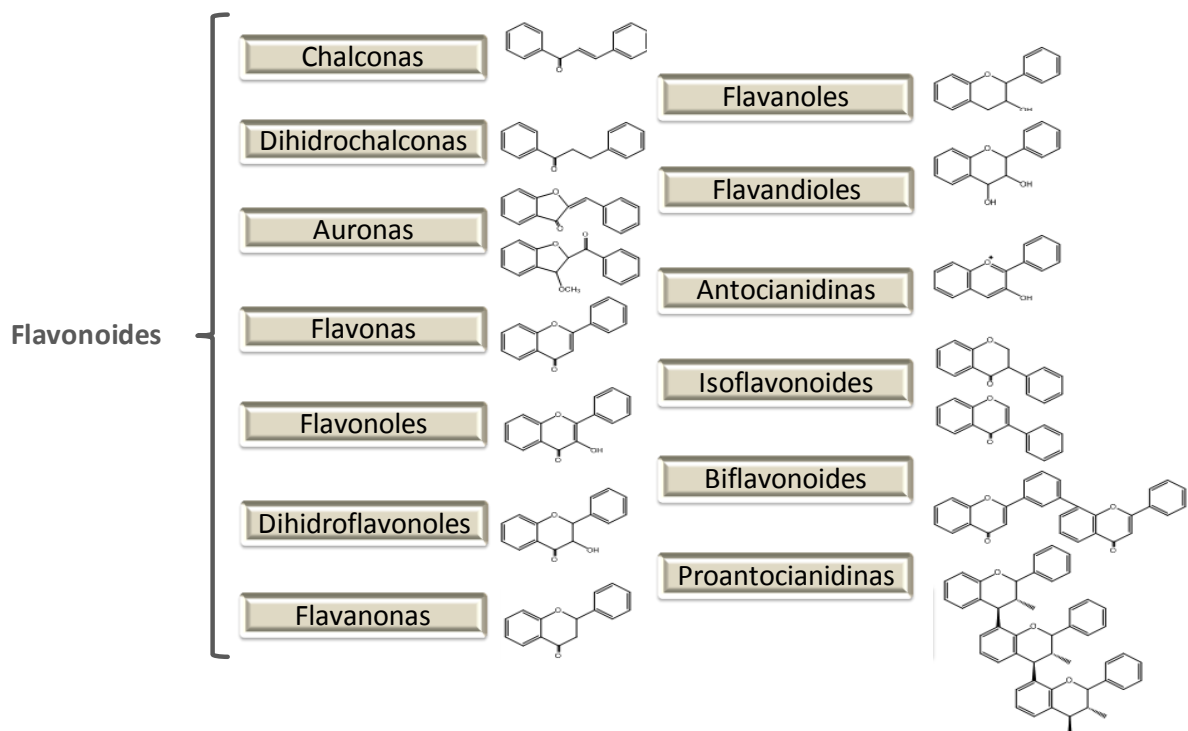


Figura 4. Clasificación y estructura de los flavonoides.

3. BIOACTIVIDAD DE TERPENOS Y COMPUESTOS FENÓLICOS.

En este apartado, se exponen las propiedades biológicas atribuidas a los compuestos descritos en el apartado anterior, los terpenos y los compuestos fenólicos.

Los compuestos de naturaleza terpénica poseen propiedades bioactivas tales como actividad bactericida, virucida, fungicida, antiparasitaria e insecticida^{21,22}. De ahí, que los aceites esenciales de plantas se hayan utilizado ampliamente en aplicaciones medicinales y cosméticas desde la Antigüedad. Hoy en día, se utilizan especialmente en conservación de los alimentos y también como

²¹ Tepe, B. *et al.*, 2005. Antimicrobial and antioxidant activities of the essential oil and various extracts of *Salvia tomentosa* Miller (Lamiaceae). *Food Chemistry*, 90(3), pp.333–340.

²² Karaman, S. *et al.*, 2001. Antibacterial and antifungal activity of the essential oils of *Thymus revolutus* Celak from Turkey. *Journal of Ethnopharmacology*, 76(2), pp.183–186.

antimicrobianos, analgésicos, sedantes, remedios anti-inflamatorios, espasmolíticos y anestésicos locales²³⁻²⁵.

En cuanto a los compuestos fenólicos, numerosos estudios han demostrado que poseen una amplia gama de propiedades bioactivas. Entre ellas, destacan la actividad antioxidante, anticancerígena, antimicrobiana, antiinflamatoria, hipocolesterolemia, etc²⁶⁻²⁹.

A continuación, se explica de manera detallada las propiedades antioxidante, antimicrobiana y anticancerígena, dado que en la presente memoria se incluyen ensayos de valoración de tales actividades biológicas.

3.1. Actividad antioxidante.

Existen evidencias científicas sobre la relación entre el estrés oxidativo y la aparición de enfermedades crónicas, como por ejemplo, las enfermedades cardiovasculares, el Alzheimer, la artritis reumatoide, etc³⁰⁻³². El estrés oxidativo es un estado del organismo causado por el desequilibrio entre la producción de radicales libres y la capacidad de detoxificar rápidamente los reactivos

²³ Guimarães, R., Sousa, M.J. & Ferreira, I.C.F.R., 2010. Contribution of essential oils and phenolics to the antioxidant properties of aromatic plants. *Industrial Crops and Products*, 32(2), pp.152–156.

²⁴ Bakkali, F. *et al.*, 2008. Biological effects of essential oils-a review. *Food and Chemical Toxicology*, 46(2), pp.446–475.

²⁵ Holley, R. a. & Patel, D., 2005. Improvement in shelf-life and safety of perishable foods by plant essential oils and smoke antimicrobials. *Food Microbiology*, 22(4), pp.273–292.

²⁶ Fernández-Panchón, M.S. *et al.*, 2008. Antioxidant activity of phenolic compounds: from in vitro results to in vivo evidence. *Critical Reviews in Food Science and Nutrition*, 48(7), pp.649–671.

²⁷ García-Lafuente, A. *et al.*, 2009. Flavonoids as anti-inflammatory agents: implications in cancer and cardiovascular disease. *Inflammation Research*, 58(9), pp.537–552.

²⁸ Ogier, N. *et al.*, 2013. LDL-cholesterol-lowering effect of a dietary supplement with plant extracts in subjects with moderate hypercholesterolemia. *European Journal of Nutrition*, 52(2), pp.547–557.

²⁹ Lozano-Sánchez, J. *et al.*, 2010. Prediction of extra virgin olive oil varieties through their phenolic profile. Potential cytotoxic activity against human breast cancer cells. *Journal of Agricultural and Food Chemistry*, 58(18), pp.9942–9955.

³⁰ Dhalla, N.S., Temsah, R.M. & Netticadan, T., 2000. Role of oxidative stress in cardiovascular diseases. *Journal of Hypertension*, 18(6), pp. 655-73.

³¹ Zafrilla, P., Mulero, J., Xandri, J.M., Santo, E., Caravaca, G., Morillas, J.M., 2006. Oxidative stress in Alzheimer patients in different stages of the disease. *Current Medicinal Chemistry*, 13(9), pp. 1075-83.

³² Feijóo, M. *et al.*, 2010. Biomarcadores de estrés oxidativo como indicadores de actividad en la enfermedad articular inflamatoria crónica. *Reumatología Clínica*, 6(2), pp. 91–94.



intermedios o reparar el daño resultante. Los radicales libres, al presentar un electrón desapareado en el orbital más externo de su estructura atómica, son moléculas altamente reactivas e inestables^{33,34}. Para estabilizarse, capturan electrones de otras sustancias formando así, otro radical libre y dando lugar a una reacción en cadena que dañará a todos los componentes de la célula, incluyendo las proteínas, los lípidos y el ADN.

El ser humano posee un sistema antioxidante endógeno formado por enzimas como la catalasa y la superóxidodismutasa, que son capaces de proteger a las células del daño oxidativo. Sin embargo, ciertas situaciones de sobreproducción de radicales libres como el tabaquismo o la contaminación ambiental, entre otras, pueden hacer que las defensas endógenas sean insuficientes (ver Figura 5). En estos casos, la protección celular va a depender de la ingesta de antioxidantes procedentes de la dieta, antioxidantes exógenos como vitaminas, algunos minerales, carotenoides y compuestos fenólicos³⁵⁻³⁷.

³³ Venereo Gutiérrez, J.R., 2002. Daño oxidativo, radicales libres y antioxidantes. *Revista Cubana de Medicina Militar*, 31(2), pp.126–133.

³⁴ Finkel, T. & Holbrook, N.J., 2000. Oxidants, oxidative stress and the biology of ageing. *Nature*, 408, pp.239–247.

³⁵ Matés, J.M., Pérez-Gómez, C. & Núñez de Castro, I., 1999. Antioxidant enzymes and human diseases. *Clinical Biochemistry*, 32(8), pp.595–603.

³⁶ Rodríguez, C. *et al.*, 2004. Regulation of antioxidant enzymes: a significant role for melatonin. *Journal of Pineal Research*, 36(1), pp.1–9.

³⁷ Valko, M. *et al.*, 2004. Role of oxygen radicals in DNA damage and cancer incidence. *Molecular and Cellular Biochemistry*, 266, pp.37–56.





Figura 5. Causas del estrés oxidativo, situaciones de sobreproducción de radicales libres.

Es importante conocer la capacidad antioxidante de los alimentos como un primer acercamiento para evaluar sus efectos biológicos. Los mecanismos de acción en los que se basan los métodos de capacidad antioxidante pueden ser de transferencia de átomos de hidrógeno (hydrogen atom transfer) (HAT), en los que el antioxidante estabiliza los radicales libres mediante la transferencia directa de un átomo de hidrógeno, o pueden ser mecanismos de transferencia de un electrón desapareado (single electron transfer) (SET), que implica la cesión de un electrón al radical. Otra forma de acción antioxidante es la quelación de metales de transición.

En la actualidad, no existe un único método que refleje de forma completa la capacidad antioxidante de una muestra debido a los complejos procesos de

oxidación. Por un lado, los radicales libres tienen varios mecanismos de acción y por otro, en un mismo alimento pueden existir diferentes compuestos antioxidantes que actúen de forma sinérgica o inhibitoria entre sí. Por tanto, se hace necesaria la utilización de diferentes métodos de determinación³⁸.

Los métodos de determinación de capacidad antioxidante se basan en la generación de radicales libres. Son métodos espectrométricos en los que se mide la señal radiante que producen los radicales libres a una determinada longitud de onda. De esta forma, cuanto mayor sea la capacidad antioxidante de la muestra, mayor será la inhibición de radicales y, por tanto, la señal se verá disminuida. Éste es el caso de las medidas de capacidad antioxidante directas, donde la adición del antioxidante de la muestra, antes o después de la generación del radical, provoca una disminución de la señal³⁹. Entre estos métodos, se encuentran el TEAC, el DPPH, etc. También hay métodos de determinación indirecta, en los que la señal no depende directamente de la formación de los radicales libres, sino de un reactivo intermedio. Los radicales libres generados producen la pérdida o aparición del reactivo, de modo que en presencia de los antioxidantes de la muestra, se originará un aumento o disminución de la señal. Entre estos métodos, se encuentran el FRAP, el ORAC y el TBARS, entre otros⁴⁰.

En la presente tesis, se han utilizado los métodos TEAC, FRAP y ORAC, cuya fundamentación se explica más detalladamente a continuación.

³⁸ López-Alarcón, C. & Denicola, A., 2013. Evaluating the antioxidant capacity of natural products: a review on chemical and cellular-based assays. *Analytica Chimica Acta*, 763, pp.1–10.

³⁹ Huang, D. *et al.*, 2002. High-throughput assay of oxygen radical absorbance capacity (ORAC) using a multichannel liquid handling system coupled with a microplate fluorescence reader in 96-well format. *Journal of Agricultural and Food Chemistry*, 50(16), pp.4437–4444.

⁴⁰ Roginsky, V. & Lissi, E. Review of methods to determine chain-breaking antioxidant activity in food. *Roginsky, V. & Lissi, E., 2005. Review of methods to determine chain-breaking antioxidant activity in food. Food Chemistry*, 92(2), pp.235–254.



- **TEAC (trolox equivalent antioxidant capacity).**

A partir de la oxidación del ácido 2,2'-azinobis (3-etilbenzotiazolín)-6-sulfónico (ABTS) (Figura 6) se genera el radical $ABTS^{\bullet+}$. Éste se puede obtener a través de una reacción química (dióxido de manganeso, persulfato potásico)^{41, 42}, enzimática (peroxidasa, mioglobulina) o también electroquímica⁴³. La forma más usual es utilizando persulfato potásico. El radical $ABTS^{\bullet+}$ obtenido es de color verde-azulado, estable y con un espectro de absorción ultravioleta-visible (UV-vis). Con este radical se pueden medir tanto compuestos de naturaleza hidrofílica como lipofílica⁴⁴.

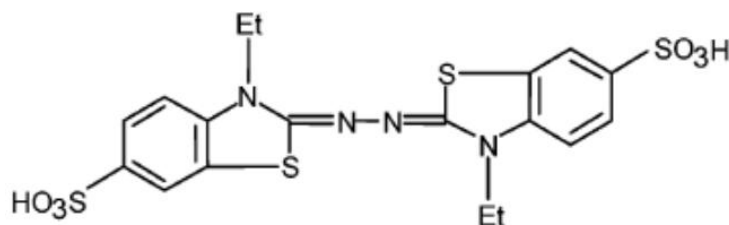


Figura 6. Estructura del ácido 2,2'-azinobis (3-etilbenzotiazolín)-6-sulfónico (ABTS).

En presencia de los antioxidantes de la muestra, se inhibe la formación de radical $ABTS^{\bullet+}$, disminuyendo por tanto la señal de absorbancia medida.

- **FRAP (ferric ion reducing antioxidant power).**

Este método se basa en la capacidad de una sustancia antioxidante para reducir el Fe^{3+} a Fe^{2+} , así el complejo férrico-2, 4,6-tripiridil-s-triazina (TPTZ) incoloro es reducido al complejo ferroso que presenta color azul (Figura 7).

⁴¹ Sellappan, S., Akoh, C.C. & Krewer, G., 2002. Phenolic compounds and antioxidant capacity of Georgia-grown blueberries and blackberries. *Journal of Agricultural and Food Chemistry*, 50(8), pp.2432–2438.

⁴² Arena, E., Fallico, B. & Maccarone, E., 2001. Evaluation of antioxidant capacity of blood orange juices as influenced by constituents, concentration process and storage. *Food Chemistry*, 74, pp.423–427.

⁴³ Imeh, U. & Khokhar, S., 2002. Distribution of conjugated and free phenols in fruits: antioxidant activity and cultivar variations. *Journal of Agricultural and Food Chemistry*, 50(22), pp.6301–6306.

⁴⁴ Antolovich, M. *et al.*, 2002. Methods for testing antioxidant activity. *The Analyst*, 127(1), pp.183–198.

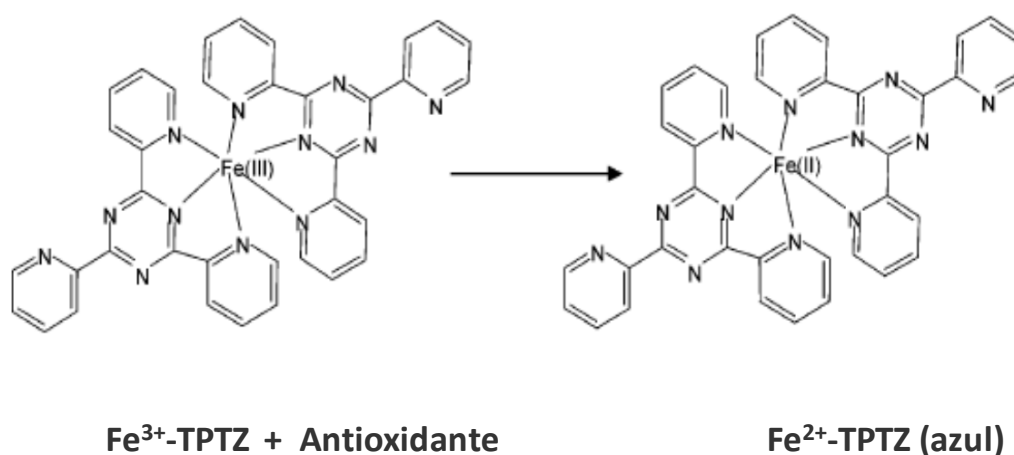


Figura 7. Reacción de reducción del TPTZ producida en el método FRAP.

Se mide la absorbancia del Fe²⁺, de modo que al añadir la muestra, a mayor capacidad reductora, mayor es la concentración de Fe²⁺ y por tanto, más alta será la señal de absorbancia⁴⁵.

- **ORAC (oxigen radical absorbance capacity).**

Se basa en la capacidad de los compuestos antioxidantes de donar átomos de hidrógeno a los radicales libres y bloquearlos. En este método se genera el radical peroxilo a partir de una molécula orgánica, el AAPH (2,2'-Azobis-(2-aminopropano)-dihidrocloruro). Los radicales oxidan la fluoresceína, por lo que ésta deja de emitir fluorescencia (Figura 8)⁴⁶.

⁴⁵ Benzie, I.F. & Strain, J.J., 1996. The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay. *Analytical Biochemistry*, 239(1), pp.70–76.

⁴⁶ Ou, B., Hampsch-Woodill, M. & Prior, R.L., 2001. Development and validation of an improved oxygen radical absorbance capacity assay using fluorescein as the fluorescent probe. *Journal of Agricultural and Food Chemistry*, 49(10), pp.4619–4626.

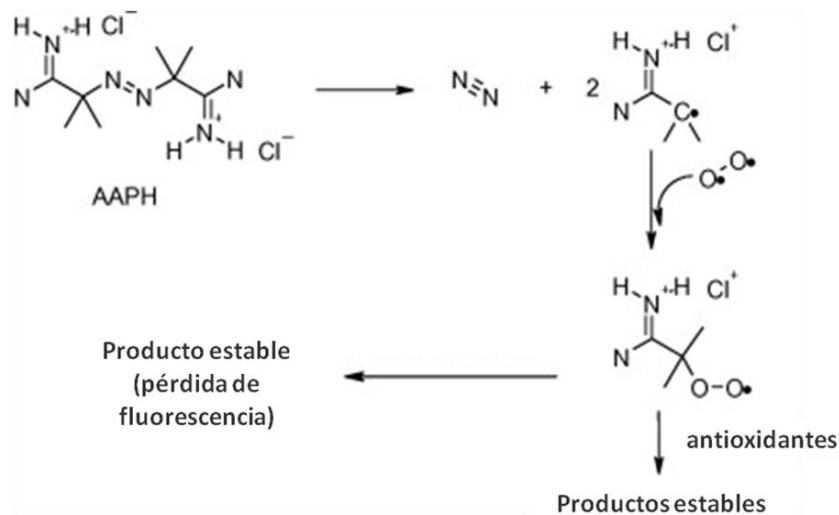


Figura 8. Reacción de oxidación de la fluoresceína por el AAPH que se produce en el método ORAC.

En presencia de los antioxidantes de la muestra, se inhibe la oxidación de la fluoresceína, siendo menor el decaimiento de la señal de fluorescencia.

3.2. Actividad antimicrobiana.

Con respecto a la actividad antimicrobiana, definimos los antimicrobianos como sustancias que evitan el crecimiento o destruyen los microorganismos colonizadores del organismo, produciendo ninguna o muy baja toxicidad sobre éste. Cuando el organismo es colonizado por especies patógenas, se produce un proceso infeccioso que depende de factores como la ruta de entrada del agente patógeno, la carga bacteriana, el periodo de incubación, la virulencia intrínseca y toxicidad del patógeno, el poder invasivo y el estado inmune del huésped. Las bacterias infecciosas se reproducen rápidamente en el organismo huésped excretando toxinas que dañan los tejidos y produciendo una afluencia de

glóbulos blancos (neutrófilos) que fagocitarán a los microorganismos agresores, comenzando así un proceso inflamatorio asociado a la infección^{47,48}.

Las bacterias pueden englobarse en Gram negativas y Gram positivas. Las bacterias Gram negativas como *Escherichia coli* no se tiñen de azul en la tinción de Gram debido a la organización de su envuelta celular, con una membrana lipídica externa recubierta de lipopolisacáridos y una membrana lipídica interna envolviendo a una fina pared de peptidoglicanos. Por el contrario, las bacterias Gram positivas como *Staphylococcus aureus* poseen una única membrana lipídica y una gruesa pared de peptidoglicanos, tiñéndose de azul en la tinción de Gram (Figura 9).

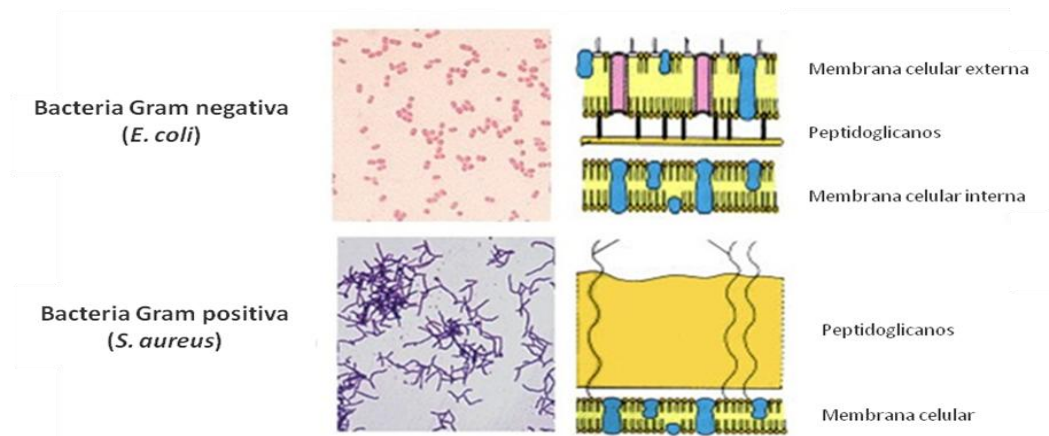


Figura 9. Tinción de Gram (izquierda) y esquema simplificado de la pared celular bacteriana (derecha). Modelo de bacteria Gram negativa (superior) y Gram positiva (inferior.)

Los extractos de plantas se han convertido en potenciales alternativas para la industria farmacológica debido a la elevada actividad antimicrobiana de los compuestos naturales y al desarrollo de multirresistencia por parte de los

⁴⁷ Daglia, M., 2012. Polyphenols as antimicrobial agents. *Current Opinion in Biotechnology*, 23(2), pp.174–181.

⁴⁸ Taguri, T., Tanaka, T. & Kouno, I., 2004. Antimicrobial activity of 10 different plant polyphenols against bacteria causing food-borne disease. *Biological & Pharmaceutical Bulletin*, 27(12), pp.1965–1969.

microorganismos a los antibióticos de uso común. Además, la utilización de antimicrobianos naturales e inocuos contribuye a disminuir la liberación de sustancias potencialmente tóxicas y xenobióticas al medio ambiente. Por ello, los antimicrobianos procedentes de plantas son de gran interés para la producción de cosméticos, jabones, colonias, perfumes y maquillaje⁴⁹.

Entre los compuestos con capacidad antimicrobiana, se han de destacar los compuestos de naturaleza terpénica como por ejemplo, el timol del tomillo o el cinamaldehído de la canela⁵⁰.

También existen evidencias sobre la capacidad antimicrobiana de los polifenoles, donde destacan aquellos que contienen un grupo galoil en su estructura, como es el caso de la epigalocatequina galato. Es importante añadir que, por lo general, el efecto antimicrobiano de los polifenoles es superior en bacterias Gram positivas, ya que los lipopolisacáridos que recubren la membrana lipídica externa de las bacterias Gram negativas impiden su permeabilización^{51,23}.

Es muy útil conocer la capacidad que tiene un extracto vegetal de inhibir el crecimiento bacteriano para determinar el posible uso que se le dará. El fundamento del ensayo de valoración de la capacidad antimicrobiana de un extracto vegetal consiste en incubar las bacterias en un medio de cultivo en presencia de diferentes concentraciones del extracto. Como colorante se usa el 2-(p-iodophenyl)-3-(p-nitrophenyl)5-phenyltetrazolium chloride (INT), una sal de tetrazolium que se reduce rápidamente por las deshidrogenasas de la cadena de

⁴⁹ Aburjai, T. & Natsheh, F.M., 2003. Plants used in cosmetics. *Phytotherapy Research*, 17(9), pp.987–1000.

⁵⁰ Prabuseenivasan, S., Jayakumar, M. & Ignacimuthu, S., 2006. In vitro antibacterial activity of some plant essential oils. *BMC Complementary and Alternative Medicine*, 6, pp.39–46.

⁵¹ Barrajón-Catalán, E. *et al.*, 2010. *Cistaceae* aqueous extracts containing ellagitannins show antioxidant and antimicrobial capacity, and cytotoxic activity against human cancer cells. *Food and Chemical Toxicology*, 48(8-9), pp.2273–2282.



transporte de electrones asociada a la respiración de las bacterias. Estas sales, al reducirse, precipitan formando un complejo insoluble en agua de color rojo intenso medible espectrofotométricamente a 570 nm. Midiendo la absorbancia a esta longitud de onda se puede determinar el porcentaje de células que siguen vivas, y por consiguiente, las que han muerto, teniendo controles positivos (donde las bacterias crecen hasta su máximo) y negativos (donde no hay crecimiento bacteriano debido al uso de un antibiótico). Como control del crecimiento bacteriano se usa la medida de la absorbancia a 620 nm. Suele utilizarse un modelo de bacteria Gram positiva y otro de Gram negativa y los resultados de este ensayo se expresan como la concentración mínima inhibitoria de extracto requerida para inhibir el 50% del crecimiento bacteriano (MIC_{50})⁵².

3.3. Actividad anticancerígena.

En la búsqueda de compuestos bioactivos, no pueden obviarse aquellos compuestos con actividad anticancerígena.

Dado que las células cancerosas se caracterizan por la proliferación celular descontrolada, cualquier molécula capaz de inhibir la proliferación de las células cancerosas puede ser útil como potencial agente quimio-preventivo.

Para analizar *in vitro* la actividad anticancerígena de extractos vegetales, suelen utilizarse ensayos de viabilidad celular que indican la citotoxicidad del tratamiento al que se someten las células con dichos extractos. Dentro de los ensayos colorimétricos, uno de los más usados es el MTT (bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolium), una sal de tetrazolio que al reducirse mediante la succinato deshidrogenasa de las mitocondrias funcionales da lugar

⁵² Caviedes, L., Delgado, J. & Gilman, R.H., 2002. Tetrazolium microplate assay as a rapid and inexpensive colorimetric method for determination of antibiotic susceptibility of mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 40(5), pp.1873–1874.



al formazán, de color azul intenso. Dicha coloración nos sirve para discernir entre las células resistentes y lábiles frente al tratamiento con los extractos. Los resultados de este ensayo pueden expresarse como la concentración de extracto a la cual muere el 50% de las células (IC_{50}).

Especial interés para la presente tesis doctoral adquiere el **cáncer de mama**, siendo el tipo de cáncer más frecuente en las mujeres occidentales. Representa alrededor del 30% de todos los tumores, y aproximadamente el 20% de todas las muertes relacionadas con esta enfermedad.

Los tumores originados a partir de tejido mamario son dependientes de estrógenos para su proliferación⁵³⁻⁵⁵.

El receptor de estrógenos (RE) es el principal encargado de recibir los estímulos estrogénicos y convertirlos en una respuesta que implique la proliferación de las células. Por tanto, en cualquier alteración de la respuesta a los estímulos estrogénicos, forzosamente debe estar implicado el receptor de estrógenos. Por ello, la terapia antiestrogénica suele ser efectiva en la mayoría de los casos. Las células RE positivas pueden tratarse con antiestrogénicos (como el Tamoxifeno), siendo estos tumores los que presentan un mejor pronóstico. Sin embargo, las células cancerosas pueden sufrir transformaciones adicionales, de manera que se pierde dicha hormonodependencia. En este estadio, los tumores presentan un peor pronóstico y se requiere de tratamientos más agresivos como quimioterapia y radioterapia.

⁵³ Perry, R.R., Kang, Y. & Greaves, B., 1995. Effects of tamoxifen on growth and apoptosis of estrogen-dependent and -independent human breast cancer cells. *Annals of Surgical Oncology*, 2(3), pp.238–245.

⁵⁴ Schrader, W.. *et al.*, 1974. Molecular mechanisms of steroid hormone action. *Journal of Steroid Biochemistry*, 5, pp.989–999.

⁵⁵ Clarke, R., Dickson, R.B. & Brünner, N., 1990. The process of malignant progression in human breast cancer. *Annals of Oncology*, 1(6), pp. 401-407.



Otro receptor que aparece mutado en aproximadamente el 30 % de los casos es el receptor ErbB2 (HER2), asociado a la activación de la proliferación celular. El cáncer de mama HER2 positivo se trata actualmente con anticuerpos anti-HER2 (Trastuzumab), aunque el pronóstico de este tipo de cánceres suele ser peor⁵⁶. Al igual que los tumores RE positivos pueden perder su hormonodependencia, los tumores HER2 positivos pueden insensibilizarse al tratamiento con anticuerpos, siendo, en este último caso, tumores con un tratamiento especialmente complicado y desfavorable. Es por esto último que, cada vez más, se buscan tratamientos alternativos para el cáncer de mama. En el marco de la investigación de compuestos con actividad anticancerígena frente al cáncer de mama, numerosos estudios se han basado en los polifenoles del vino, el aceite de oliva, la granada u otros extractos vegetales como la curcumina o la genisteína de la soja⁵⁷⁻⁶³.

⁵⁶ Collingwood, T.N., Urnov, F.D. & Wolffe, a P., 1999. Nuclear receptors: coactivators, corepressors and chromatin remodeling in the control of transcription. *Journal of Molecular Endocrinology*, 23(3), pp.255–275.

⁵⁷ Kavanagh, K.T. *et al.*, 2001. Green tea extracts decrease carcinogen-induced mammary tumor burden in rats and rate of breast cancer cell proliferation in culture. *Journal of Cellular Biochemistry*, 82(3), pp.387–398.

⁵⁸ Thangapazham, R.L. *et al.*, 2007. Green tea polyphenols and its constituent epigallocatechin gallate inhibits proliferation of human breast cancer cells in vitro and in vivo. *Cancer Letters*, 245(1-2), pp.232–241.

⁵⁹ Liao, S. *et al.*, 1995. Growth inhibition and regression of human prostate and breast tumors in athymic mice by tea epigallocatechin gallate. *Cancer Letters*, 96(2), pp.239–243.

⁶⁰ Damianaki, A. *et al.*, 2000. Potent inhibitory action of red wine polyphenols on human breast cancer cells. *Journal of Cellular Biochemistry*, 78(3), pp.429–441.

⁶¹ Hakimuddin, F., Paliyath, G. & Meckling, K., 2006. Treatment of mcf-7 breast cancer cells with a red grape wine polyphenol fraction results in disruption of calcium homeostasis and cell cycle arrest causing selective cytotoxicity. *Journal of Agricultural and Food Chemistry*, 54(20), pp.7912–7923.

⁶² Kim, N.D. *et al.*, 2002. Chemopreventive and adjuvant therapeutic potential of pomegranate (*Punica granatum*) for human breast cancer. *Breast Cancer Research and Treatment*, 71(3), pp.203–217.

⁶³ Owen, R.W. *et al.*, 2000. The antioxidant/anticancer potential of phenolic compounds isolated from olive oil. *European Journal of Cancer*, 36(10), pp.1235–1247.



4. EXTRACCIÓN DE COMPUESTOS BIOACTIVOS EN MUESTRAS VEGETALES.

Las matrices vegetales son, en general, muy complejas por lo que normalmente no es posible abordar el análisis directo de la matriz en su conjunto. Es necesario realizar un tratamiento de muestra adecuado, capaz de eliminar todos los interferentes presentes en la matriz y asegurar la obtención de un extracto de la muestra compatible con la técnica analítica que se vaya a utilizar para la determinación. El tratamiento de la muestra constituye una etapa crucial del proceso analítico ya que de él depende en gran medida la calidad y reproducibilidad de los resultados analíticos. Generalmente, consta de una etapa de homogeneización y secado, seguida del proceso de extracción de los componentes de interés. También puede ser necesario realizar la etapa de secado después de la extracción, asegurando la conservación prolongada del extracto previa al análisis⁶⁴.

A continuación, se describe el fundamento de aquellos sistemas de extracción y secado utilizados en la presente tesis.

En la Figura 10 se presentan las metodologías de extracción clasificadas como convencionales y alternativas.

⁶⁴ Escarpa, a. & González, M.C., 2001. An overview of analytical chemistry of phenolic compounds in foods. *Critical Reviews in Analytical Chemistry*, 31(2), pp.57–139.



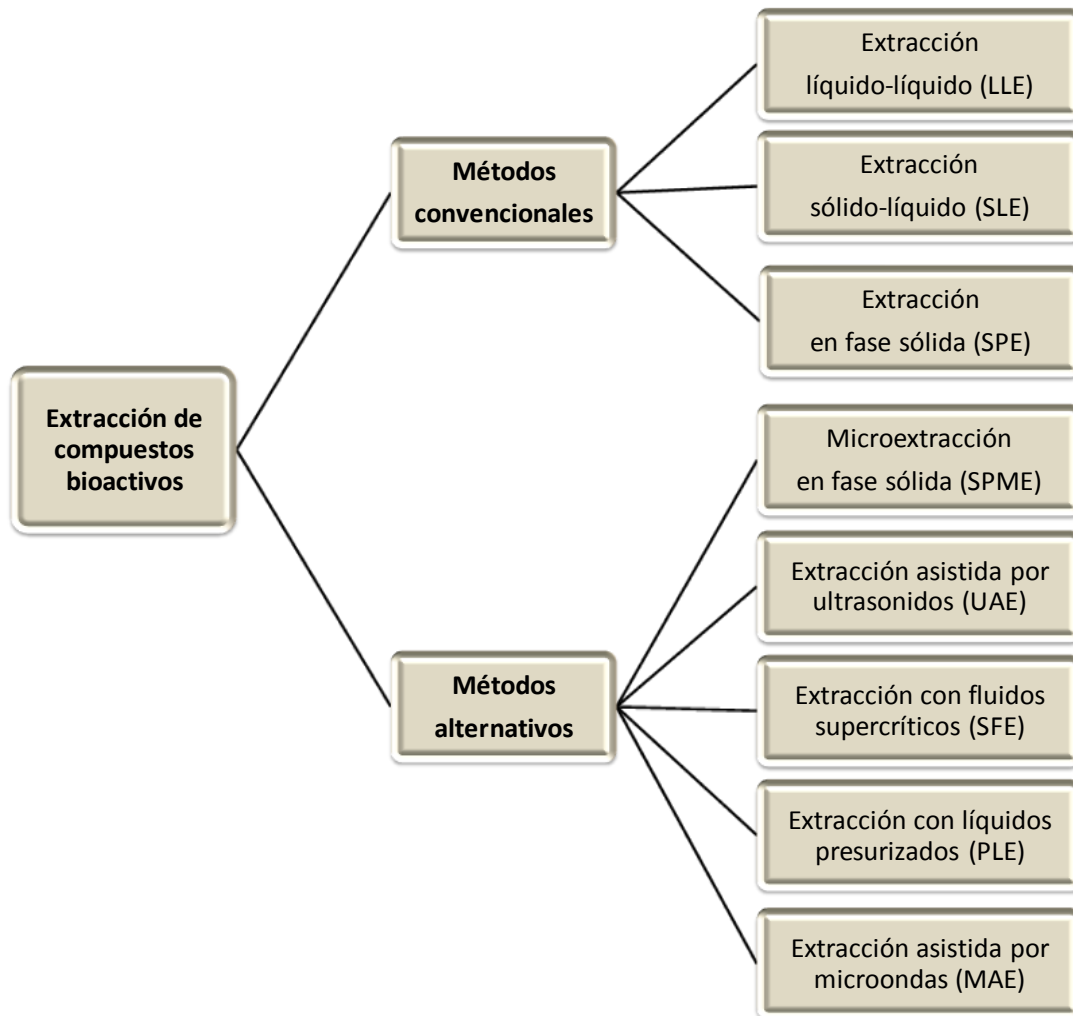


Figura 10. Métodos de extracción de compuestos bioactivos.

Se han descrito en bibliografía muchos procedimientos para la extracción de compuestos bioactivos de matrices vegetales, en su mayoría utilizando **extracción líquido-líquido (LLE)** y **extracción sólido-líquido (SLE)**⁶⁵. En ambos casos, se trata de extraer el analito (o grupo de compuestos) empleando un disolvente adecuado en el que es insoluble el resto de la matriz. La extracción líquido-líquido se utiliza para matrices líquidas, mientras que en la extracción

⁶⁵ Harnly, J.M., Bhagwat, S. & Lin, L.-Z., 2007. Profiling methods for the determination of phenolic compounds in foods and dietary supplements. *Analytical and Bioanalytical Chemistry*, 389(1), pp.47–61.

sólido-líquido, el analito forma parte de una matriz sólida. En este tipo de extracciones juega un papel decisivo la solubilidad de los compuestos bioactivos en los disolventes utilizados, la cual está condicionada por la naturaleza química del compuesto. Se trata de un proceso que implica dos etapas: puesta en contacto de la matriz con el disolvente de extracción y separación de ambas fases tras el tiempo requerido para obtener una máxima recuperación del analito de interés⁶⁶.

La extracción sólido-líquido es la más ampliamente utilizada en muestras vegetales por ser un método relativamente económico y sencillo. Para extraer compuestos fenólicos, suelen utilizarse disolventes orgánicos tales como metanol o etanol⁶⁷. En el caso de compuestos volátiles tales como los terpenos es más adecuado el uso de disolventes como pentano, hexano, benceno, acetona, acetato de etilo, acetato de metilo o éter dietílico. Si lo que se desea es obtener el aceite esencial de la planta, se suelen utilizar procesos de destilación (hidrodestilación o por arrastre de vapor). Este método requiere de la concentración posterior, seguida de la evaporación del disolvente, parte de la operación en la cual deben tomarse precauciones para evitar pérdidas por evaporación de compuestos volátiles bioactivos⁶⁸.

En comparación con la extracción líquido-líquido, la **extracción en fase sólida** (SPE) es una separación más completa donde no sólo se pueden utilizar disolventes inmiscibles, sino que no hay apenas restricciones en la elección de los mismos. Además, evita la posible formación de emulsiones y no suele requerir tratamiento de calor. La extracción SPE puede llevarse a cabo de forma

⁶⁶ García-Salas, P. *et al.*, 2010. Phenolic-compound-extraction systems for fruit and vegetable samples. *Molecules*, 15(12), pp.8813–8826.

⁶⁷ Prati, S. *et al.*, 2007. Composition and content of seed flavonoids in forage and grain legume crops. *Journal of Separation Science*, 30(4), pp.491–501.

⁶⁸ Péres, V.F. *et al.*, 2006. Comparison of soxhlet, ultrasound-assisted and pressurized liquid extraction of terpenes, fatty acids and Vitamin E from *Piper gaudichaudianum* Kunth. *Journal of Chromatography. A*, 1105(1-2), pp.115–118.



manual o automatizada. Los diferentes formatos en los que se presenta la fase estacionaria son discos, placas y cartuchos. Los dos primeros son utilizados para volúmenes muy pequeños tanto de muestra como de elución (entre 0.5 y 2.5 mL). Los cartuchos son los más utilizados en la extracción de compuestos bioactivos de matrices alimentarias líquidas. En cuanto a las diferentes tipologías de rellenos, se encuentran las fases poliméricas y las fases en base sílica, a partir de las cuales, se pueden desarrollar una gran variedad de fases estacionarias diferentes en función de la modificación o adición de determinados grupos funcionales. Las fases en base sílica son mucho más selectivas y, aunque presentan una menor estabilidad química que las fases poliméricas, son las más ampliamente utilizadas en la extracción de compuestos bioactivos⁶⁹. Normalmente, se utiliza la SPE para extraer analitos semivolátiles o no volátiles. Para la extracción de los polifenoles del aceite, también se han desarrollado diferentes métodos, siendo la extracción líquido-líquido y la extracción en fase sólida con cartuchos DIOL las que producen mejores resultados, en relación al número de compuestos extraídos y el nivel de preconcentración alcanzado⁷⁰.

De la necesidad de utilizar sistemas más sensibles, selectivos, rápidos y ecológicos que los métodos convencionales anteriormente citados, han surgido otras metodologías de extracción de compuestos bioactivos con la ventaja de requerir una cantidad menor o nula de disolventes orgánicos. Algunas de estas tecnologías alternativas son la extracción asistida por ultrasonidos (UAE), la microextracción en fase sólida (SPME), la extracción con fluidos supercríticos (SFE), con líquidos presurizados (PLE) y la extracción asistida por microondas (MAE)⁶⁶. Entre dichas tecnologías, destaca la extracción asistida por ultrasonidos

⁶⁹ Palma, M., Piñeiro, Z. & Barroso, C.G., 2002. In-line pressurized-fluid extraction-solid-phase extraction for determining phenolic compounds in grapes. *Journal of Chromatography. A*, 968(1-2), pp.1-6.

⁷⁰ Gómez-Caravaca, A.M. *et al.*, 2005. Electrophoretic identification and quantitation of compounds in the polyphenolic fraction of extra-virgin olive oil. *Electrophoresis*, 26(18), pp.3538-3551.



y la microextracción en fase sólida, dada su utilización en el desarrollo experimental de la tesis, por lo que se describen a continuación.

La **extracción por ultrasonidos** (UAE) es un proceso simple, normalmente utilizado junto con la extracción con disolventes para acelerar el proceso y limitar la degradación de los compuestos de la matriz. Se fundamenta en la utilización de sonidos de alta frecuencia que permiten aumentar el rendimiento de la extracción. Su mayor rendimiento se ha atribuido a la propagación de ondas de presión ultrasónica y el fenómeno de cavitación resultante; la presión generada produce el paso de estado líquido a vapor con la consiguiente formación de cavidades. El sistema de extracción UAE es un proceso apropiado para la extracción de compuestos fenólicos en muestras vegetales, ampliamente documentado en bibliografía. En estos casos, es conveniente reducir el tamaño de las partículas para aumentar el área de exposición al solvente y la cavitación producida^{71,72}.

La **microextracción en fase sólida** (SPME) constituye un sistema alternativo que elimina totalmente la utilización de disolventes. Es ampliamente utilizado para extraer compuestos volátiles a partir de muestras vegetales, e incluso puede utilizarse para la extracción de compuestos volátiles emitidos *in vivo* por la planta. Se basa en la utilización de una fibra de sílice fundida revestida con un material adsorbente, capaz de extraer diferentes clases de analitos presentes en diferentes medios, tanto líquidos como gases. La fibra de SPME está protegida por una jeringa que permite la penetración del *septum* del vial usado para la extracción (donde se coloca la muestra), así como su introducción en el puerto de inyección del equipo cromatográfico con el que se procederá al análisis

⁷¹ Dobiáš, P. *et al.*, 2010. Comparison of pressurised fluid and ultrasonic extraction methods for analysis of plant antioxidants and their antioxidant capacity. *Central European Journal of Chemistry*, 8(1), pp.87–95.

⁷² Vilkh, K. *et al.*, 2008. Applications and opportunities for ultrasound assisted extraction in the food industry - A review. *Innovative Food Science & Emerging Technologies*, 9(2), pp.161–169.



posterior. Para obtener resultados cuantitativos reproducibles, la fibra y la muestra deben llegar al punto de equilibrio, en el cual la cantidad de analito liberado por la fibra es proporcional a la concentración del compuesto en la muestra. El tiempo requerido para alcanzar el equilibrio depende de la volatilidad y polaridad del analito y de las propiedades del adsorbente⁷³. Por ello, es imprescindible seleccionar cuidadosamente la polaridad y el grosor del recubrimiento de la fibra. Mientras que las fibras recubiertas por una capa fina aseguran una difusión rápida y son capaces de liberar los compuestos semivolátiles, aquellas cuya capa es más gruesa retienen mejor los compuestos altamente volátiles hasta el momento de la desorción de los mismos. Cuando la extracción termina, la fibra se retira, y se introduce en el puerto de inyección del cromatógrafo de gases (GC) para la desorción de los compuestos extraídos.

En un proceso característico de SPME, se pueden desarrollar dos métodos básicos de extracción: la extracción directa, donde la fibra se expone directamente dentro de la matriz, o el modo de extracción *headspace* o de espacio de cabeza (HS-SPME), en el que se inserta la aguja que contiene la fibra a través del *septum* del vial y, empujando el émbolo, se expone la fibra a los compuestos volátiles del espacio de cabeza de la muestra en el vial. En la extracción HS-SPME, sólo se extraen efectivamente los analitos relativamente volátiles. Por ello, este modo se considera un método rápido y simple para la extracción de compuestos volátiles, ventajoso además para aquellas muestras que presentan interferencias de alto peso molecular. En comparación con los largos y tediosos procesos de obtención de aceites esenciales de las plantas, la extracción de los compuestos volátiles mediante HS-SPME y su desorción térmica directa desde la fibra evita la necesidad de utilizar disolventes que

⁷³ Górecki, T., Yu, X. & Pawliszyn, J., 1999. Theory of analyte extraction by selected porous polymer SPME fibres. *The Analyst*, 124(5), pp.643–649.



puedan contener impurezas e interferir en el análisis^{74,75}. En la Figura 11 se muestra un esquema del proceso de extracción HS-SPME, siendo en este caso un ejemplo de sistema SPME incorporado en el cromatógrafo de gases.

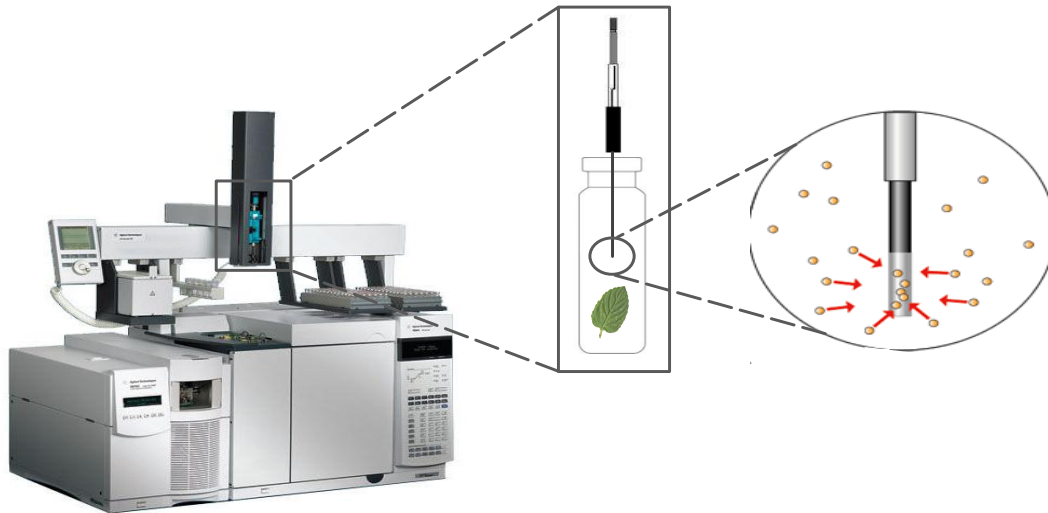


Figura 11. Sistema de microextracción en fase sólida de espacio en cabeza (HS-SPME).

Con respecto a los sistemas de secado, destacan la liofilización y el secado por *spray drying*. Brevemente, en el proceso de liofilización o *freeze drying*, se congela el producto y posteriormente se introduce en una cámara de vacío para realizar la separación del agua por sublimación, de manera que se elimina el agua desde el estado sólido al gaseoso sin pasar por el estado líquido⁷⁶. El sistema *spray drying*, también conocido como secado por aspersión o pulverización, se basa en la obtención de un producto en polvo a partir de un material líquido concentrado. Mediante pulverización, se forma una niebla que entra en contacto con una corriente de aire caliente, la cual actúa como medio

⁷⁴ Tholl, D. *et al.*, 2006. Practical approaches to plant volatile analysis. *The Plant Journal*, 45(4), pp.540–560.

⁷⁵ Dudareva, N. *et al.*, 2006. Plant volatiles: recent advances and future perspectives. *Critical Reviews in Plant Sciences*, 25(5), pp.417–440.

⁷⁶ Ratti, C., 2001. Hot air and freeze-drying of high-value foods: a review. *Journal of Food Engineering*, 49(4), pp.311–319.

calefactor y fluido de transporte^{77,78}. Otro sistema a destacar, que puede utilizarse como método de secado es la tecnología D.I.C (descompresión instantánea controlada), que consiste en un tratamiento termo-mecánico de alta temperatura y corto tiempo, combinado con una descompresión instantánea a vacío⁷⁹.

5. SEPARACIÓN DE COMPUESTOS BIOACTIVOS EN MUESTRAS VEGETALES.

Para la identificación de cada uno de los componentes bioactivos de una muestra vegetal es necesario emplear técnicas separativas. La técnica más apropiada para la separación de compuestos volátiles es la cromatografía de gases. Dicha técnica también puede utilizarse para analizar compuestos no volátiles, siendo necesario en este caso un proceso de derivatización con el fin de mejorar la volatilidad, estabilidad térmica y detección de los mismos. En el caso de los compuestos fenólicos, se pueden utilizar técnicas electroforéticas, así como cromatografía de gases. Sin embargo, la técnica más potente para la separación de estos compuestos es la cromatografía líquida de alta resolución^{80,81}.

En cualquier separación cromatográfica, los componentes de una muestra son transportados mediante una fase que se mueve en una dirección definida (fase móvil) a través de una fase que está en reposo (fase estacionaria). El contacto

⁷⁷ Vehring, R., 2008. Pharmaceutical particle engineering via spray drying. *Pharmaceutical Research*, 25(5), pp.999–1022.

⁷⁸ Gharsallaoui, A. *et al.*, 2007. Applications of spray-drying in microencapsulation of food ingredients: An overview. *Food Research International*, 40(9), pp.1107–1121.

⁷⁹ Haddad, M. Al *et al.*, 2008. Fruits & Vegetables Drying Combining Hot Air, DIC Technology and Microwaves. *International Journal of Food Engineering*, 4(6), article 9.

⁸⁰ Castro-Puyana, M. *et al.*, 2012. Recent advances in the application of capillary electromigration methods for food analysis and Foodomics. *Electrophoresis*, 33(1), pp.147–167.

⁸¹ Gómez-Caravaca, a M. *et al.*, 2006. Advances in the analysis of phenolic compounds in products derived from bees. *Journal of Pharmaceutical and Biomedical Analysis*, 41(4), pp.1220–1234.



entre las dos fases es similar a un equilibrio dinámico, en el que los componentes a separar presentarán una repartición diferencial debido a las interacciones entre la fase estacionaria y la fase móvil⁸². A continuación, se detalla de manera independiente el fundamento de las dos técnicas separativas utilizadas en la presente tesis, cromatografía de gases y cromatografía líquida.

5.1. Cromatografía de gases (GC).

En cromatografía de gases, la fase móvil es simplemente un gas portador que, por su baja densidad, no participa en el proceso de distribución del analito entre ambas fases. La fase estacionaria, contenida en la columna, es un líquido retenido o inmovilizado sobre un soporte sólido, caracterizado por unas características físicas y químicas concretas. Por ello, al fijar las condiciones de la fase estacionaria, el único parámetro que puede modificar el equilibrio en una separación GC es la temperatura.

En la Figura 12 se muestran las partes de las que consta un cromatógrafo de gases. Presenta un sistema que proporciona el gas portador, siendo los más comunes nitrógeno, helio e hidrógeno. El flujo y la presión de este gas portador están regulados por un regulador de flujo y un regulador de presión. La cantidad de muestra que entra en el sistema es controlada por un divisor de flujo, antes de la inyección de la misma. La columna se encuentra en el interior de un horno ya que los análisis se llevan a cabo a altas temperaturas para garantizar que los analitos se encuentren en fase gaseosa. Por último, se encuentra el detector, que proporciona la señal analítica⁸³.

⁸² Bélanger, J.M.R., Bissonnette, M.C. & Paré, J.R.J., 1997. Chromatography: principles and applications. In J.R.J. Paré & J.M.R. Bélanger, eds. *Instrumental Methods in Food Analysis*. pp. 1–35.

⁸³ Wang, Z. & Paré, J.R.J., 1997. Gas Chromatography (GC): principles and applications. In *Instrumental Methods in Food Analysis*. pp. 61–91.



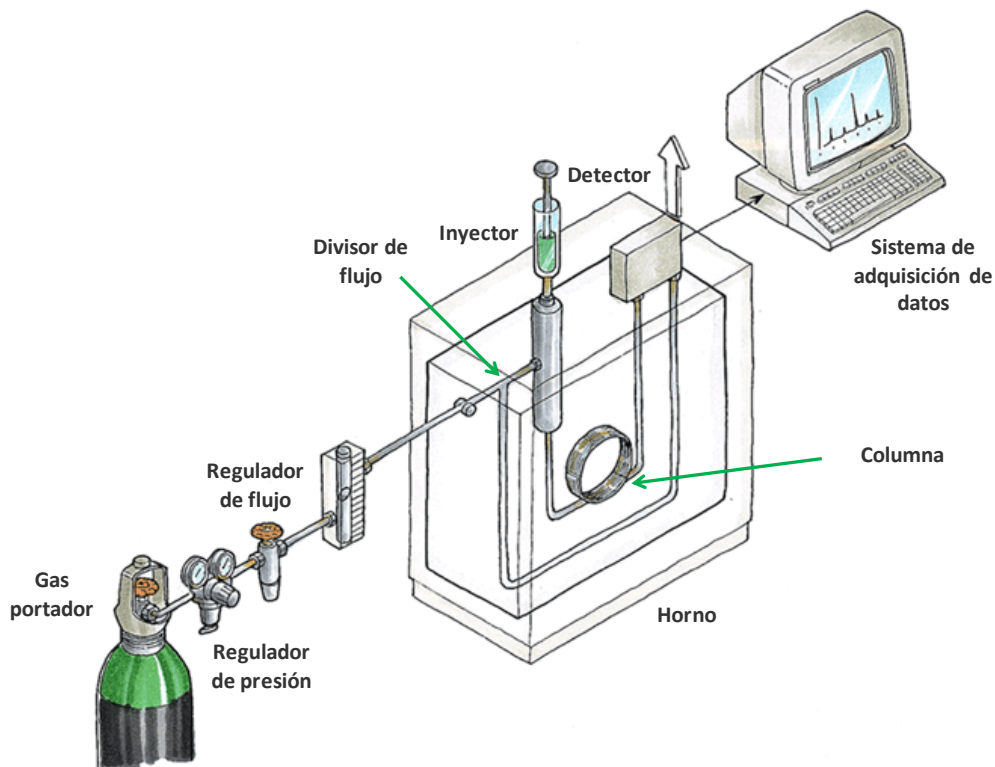


Figura 12. Esquema de las partes de un cromatógrafo de gases.

Para el análisis de compuestos volátiles por cromatografía de gases, las muestras son o bien inyectadas como extractos de disolventes en el inyector o bien, desorbidas por una fibra que se coloca directamente en el tubo de desorción térmica a una temperatura de 250-300°C. Para la separación de compuestos volátiles se utilizan comúnmente columnas capilares de sílice con diferentes fases estacionarias tales como los compuestos apolares dimetilpolisiloxanos (DB-1, DB-5, CPSil 5) y los polímeros más polares de polietilenglicol, incluyendo la Carbowax[®] 20M, la DB-Wax y la HP-20M.

Avances tecnológicos recientes han permitido el desarrollo de sistemas en continuo (on-line), los cuales combinan la extracción de los volátiles por microextracción en fase sólida (SPME) y su desorción automática, pudiendo prescindir de la SPME manual. Estos equipos cuentan con una trampa de frío o

criotrampa en la que los volátiles liberados térmicamente se concentran justo antes de la inyección en la columna cromatográfica⁸³.

Entre los equipos de cromatografía de gases disponibles en el mercado, destacamos los SPME-GC automatizados que ofrecen un avance importante, permitiendo el procesamiento rápido de múltiples viales gracias a los automuestreadores que tienen incorporados. Estos equipos constan de un brazo-robot ajustable, que permite empezar la extracción de la siguiente muestra, mientras que la anterior está siendo sometida a la separación en la columna. Por ello, los sistemas automatizados son ideales para realizar cribados de múltiples muestras de una manera rápida y automática. Otra de las diferencias con el sistema no automatizado es el sistema de agitación de la muestra. En la operación manual, se utilizan agitadores magnéticos, mientras que el equipo automatizado incorpora el sistema de agitación como una opción del automuestreador. Consiste en un pequeño motor con una leva, es decir, una pieza que unida a un eje, gira comunicando su movimiento al vial donde se encuentra la muestra. Este sistema simplifica la extracción puesto que no requiere la introducción de objetos externos a la muestra antes de su inyección, eliminando así la posibilidad de contaminación asociada a los agitadores magnéticos reusados^{74,75}.

5.2. Cromatografía líquida (LC).

La cromatografía líquida de alta resolución (HPLC) es una técnica de separación donde los componentes de la muestra se distribuyen entre una fase móvil (disolvente/s) y una fase estacionaria (relleno de la columna). La separación se produce en función de las distintas afinidades de los constituyentes de la muestra hacia cada fase. La fase estacionaria puede ser un sólido poroso, generalmente en forma particulada, o bien una fina capa de sustancia ligada a un soporte sólido, contenido en el interior de un tubo habitualmente metálico



que da lugar a la columna cromatográfica. La fase móvil es un disolvente o mezcla de ellos a los que en algunas ocasiones se les modifica el pH mediante adición de ácidos, bases o disoluciones reguladoras.

En una separación por HPLC, la fase móvil impulsada por la bomba transporta una banda de muestra a través de la columna cromatográfica. Al entrar la muestra en la columna, sus componentes interaccionan en extensión variable con la fase estacionaria y se reparten de manera distinta entre ambas fases de tal forma que aquellos que sean más afines con la fase móvil serán menos retenidos por la fase estacionaria y eluirán antes. Por el contrario, aquellos que tengan más afinidad por la fase estacionaria avanzarán más lentamente a través de la columna y eluirán más tarde. Cuando cada analito eluye del final de la columna, entra en el detector y produce de alguna forma una señal medible. La intensidad y duración de la señal estará relacionada con la cantidad o naturaleza del analito. Generalmente, la señal es amplificada y registrada por un integrador electrónico. En la Figura 13, se representa el esquema básico de un equipo de HPLC y se muestra el aspecto de uno comercial.



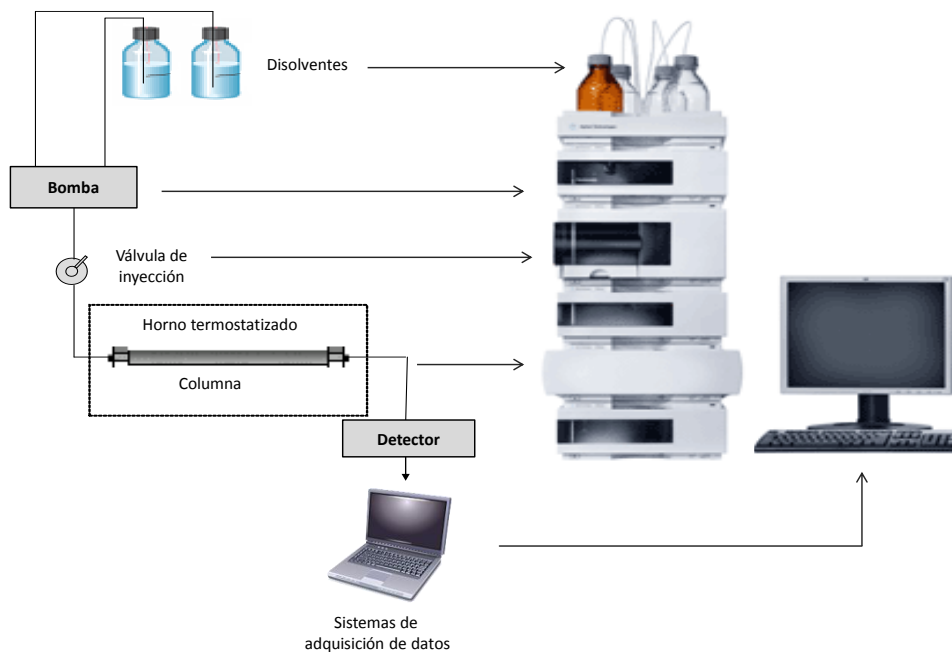


Figura 13. Esquema básico de un equipo de HPLC junto a uno comercial.

Generalmente se considera HPLC cuando la separación cromatográfica se lleva a cabo en columnas de diámetro interno de entre 1,5 y 4,5 mm. Cuando el tamaño de partícula de la fase estacionaria es inferior a 2 μm , se considera comúnmente cromatografía líquida de resolución rápida (RRLC) o cromatografía líquida de ultra alta resolución (UPLC). Estas técnicas ofrecen ventajas importantes en cuanto a resolución, rapidez y sensibilidad de la determinación analítica⁸⁴. Normalmente la longitud de la columna es menor y se usan flujos relativamente elevados, permitiendo de este modo, realizar análisis más rápidos manteniendo la eficacia constante, u obtener eficacias mayores manteniendo el mismo tiempo de análisis⁸⁵.

⁸⁴ Estella-Hermoso de Mendoza, A. *et al.*, 2009. Comparative study of A HPLC-MS assay versus an UPLC-MS/MS for anti-tumoral alkyl lysophospholipid edelfosine determination in both biological samples and in lipid nanoparticulate systems. *Journal of Chromatography B*, 877(31), pp.4035–4041.

⁸⁵ Desmet, G. *et al.*, 2006. Future of high pressure liquid chromatography: do we need porosity or do we need pressure? *Journal of Chromatography A*, 1130(1), pp.158–166.

Fundamentalmente, los compuestos fenólicos se separan mediante cromatografía de partición en fase inversa dado su carácter polar. En este caso, la fase estacionaria es un líquido apolar retenido por un soporte sólido como hidrocarburos C₈ (n-octilo) o C₁₈ (n-octadecilo) y la elución se lleva a cabo con una fase móvil de polaridad elevada, normalmente, disoluciones acuosas con cierto porcentaje de disolventes orgánicos polares.

Existe una gran diversidad de columnas disponibles comercialmente, que se diferencian, principalmente, en sus dimensiones, es decir, longitud y diámetro de la columna, así como en el tamaño de partícula de su relleno. Tradicionalmente, para el análisis de compuestos fenólicos se han empleado columnas convencionales de HPLC con un tamaño de partícula de entre 3 y 5 µm, pero la tendencia actual se centra en la disminución de dicho parámetro. En la última década, el uso de RRLC o UPLC se ha extendido ampliamente para la separación de estos compuestos en matrices vegetales⁸⁶⁻⁸⁸.

6. DETECCIÓN DE COMPUESTOS BIOACTIVOS EN MUESTRAS VEGETALES: ESPECTROMETRÍA DE MASAS (MS).

En cuanto a la gran diversidad de detectores acoplables a equipos de cromatografía, siendo el sistema de detección de espectrometría de masas el más ampliamente utilizado para la detección de compuestos bioactivos en muestras vegetales, que puede acoplarse tanto a cromatografía líquida como a cromatografía de gases. Este acoplamiento también presenta la ventaja de

⁸⁶ Cavaliere, C. *et al.*, 2008. Rapid-resolution liquid chromatography/mass spectrometry for determination and quantitation of polyphenols in grape berries. *Rapid Communications in Mass Spectrometry*, 22, pp.3089–3099.

⁸⁷ Castro-Pérez, J. *et al.*, 2005. Increasing throughput and information content for in vitro drug metabolism experiments using ultra-performance liquid chromatography coupled to a quadrupole time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, 19(6), pp.843–848.

⁸⁸ Serra, A. *et al.*, 2009. Determination of procyanidins and their metabolites in plasma samples by improved liquid chromatography-tandem mass spectrometry. *Journal of Chromatography B*, 877(11-12), pp.1169–1176.



aportar una segunda dimensión de separación ya que, tras separar los compuestos según su tiempo de retención, se produce en el espectrómetro de masas una separación en función de la relación masa/carga (m/z).

Existen diversos tipos de espectrómetros de masas, pero todos incluyen los siguientes elementos: un sistema de introducción de muestra, un medio por el que las muestras son ionizadas, un acelerador de iones mediante un campo eléctrico, un sistema de dispersión de iones de acuerdo a su relación m/z y un detector de los iones junto con un procesador adecuado de la señal (ver Figura 14)⁸⁹.

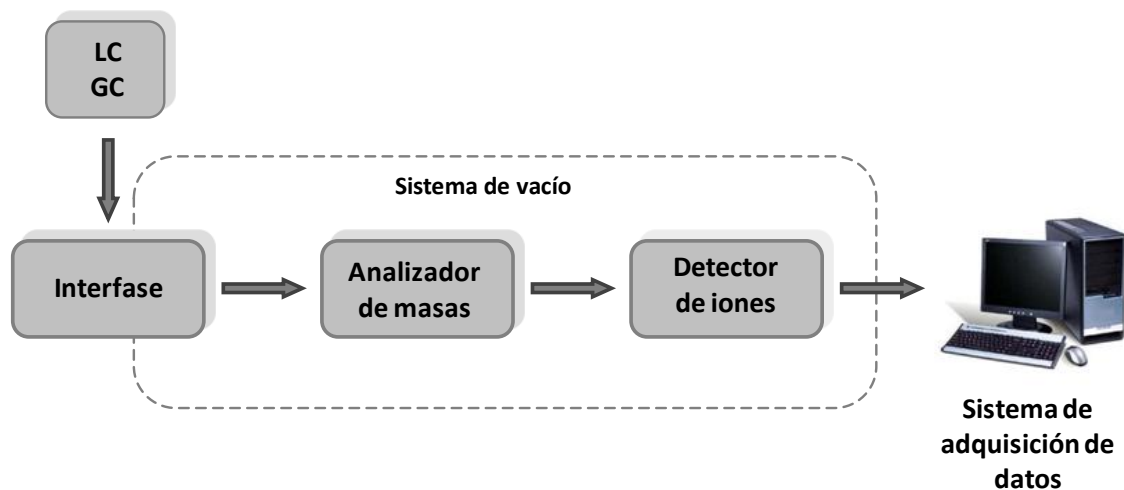


Figura 14. Esquema de las partes de un espectrómetro de masas y su acoplamiento con cromatografía líquida o de gases.

El acoplamiento entre el cromatógrafo de gases y el espectrómetro de masas se realiza mediante el uso de un **sistema de ionización** o **interfase**, siendo la de **impacto electrónico** (EI) la más popular para los análisis de volátiles de rutina y, por tanto, la que incorpora la mayoría de los cromatógrafos de gases. Consiste en un filamento cargado negativamente y calentado por una corriente que

⁸⁹ Watson, J.T. & Sparkman, O.D., 2007. *Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation*. 4 th. J. Wiley & Sons, ed., New York.

circula en el mismo, actuando como emisor de electrones. A través de un campo eléctrico, los electrones son acelerados hacia el ánodo e interaccionan con las moléculas gaseosas de la muestra, originando especies ionizadas (ver Figura 15).

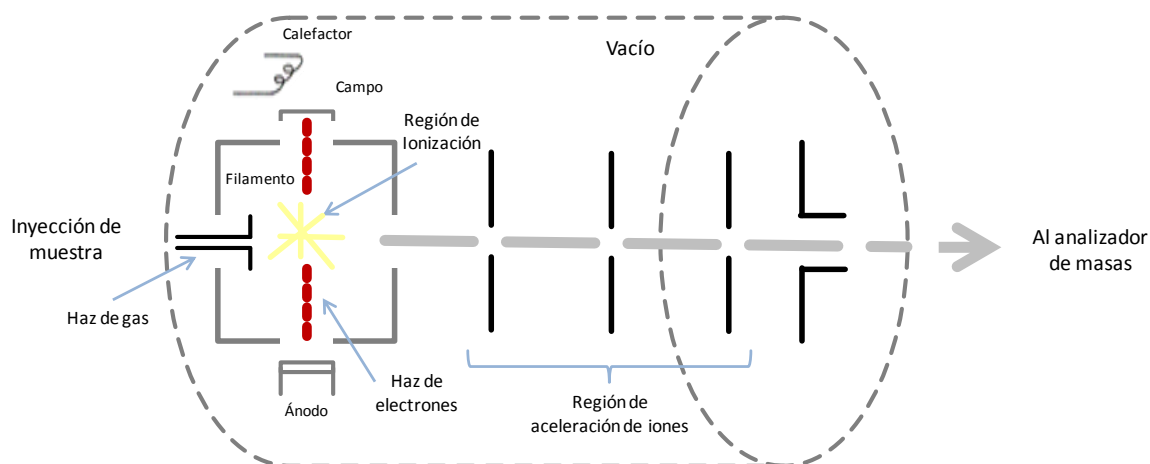


Figura 15. Esquema del sistema de ionización por impacto electrónico.

La ionización por impacto electrónico es poco eficaz, por lo que es necesario acelerar los electrones aplicando un potencial mayor de 50 V para formar un número significativo de iones. Esta gran cantidad de energía implica el estado excitado de las especies originadas, de manera que con la relajación posterior se producen rupturas de enlaces y se obtienen iones fragmentados con una relación m/z menor que la del ión molecular. Dichos fragmentos pueden a la vez, sufrir fragmentaciones. Por todo ello, se considera el impacto electrónico como un sistema de ionización duro^{90,91}.

Para llevar a cabo el acoplamiento de cromatografía líquida con espectrometría de masas, también se precisa de un sistema de ionización que asegure que las sustancias procedentes del cromatógrafo que trabaja en estado líquido entren al espectrómetro en fase gaseosa. Se han desarrollado diferentes interfases

⁹⁰ Fialkov, A.B. *et al.*, 2006. A new type of GC-MS with advanced capabilities. *International Journal of Mass Spectrometry*, 251(1), pp.47-58.

⁹¹ Hao, C., Zhao, X. & Yang, P., 2007. GC-MS and HPLC-MS analysis of bioactive pharmaceuticals and personal-care products in environmental matrices. *Trends in Analytical Chemistry*, 26(6), pp.569-580.

para el acoplamiento HPLC-MS, pero la más empleada para el análisis del perfil fenólico de muestras vegetales es el de **ionización por electrospray** (ESI). En este caso, se trata de un sistema de ionización blando que da lugar a poca fragmentación.

En el proceso de formación del *electrospray* (Figura 16), el cual se lleva a cabo a presión atmosférica, intervienen diversos mecanismos al mismo tiempo. La muestra, procedente del capilar de separación, y con la ayuda de un gas nebulizador, se carga y dispersa simultáneamente. El disolvente se va evaporando de las microgotas formadas (desolvatación) y éstas van aumentando su densidad de carga eléctrica. Como consecuencia, las gotas se encogen y los iones que se encuentran en la superficie se ven forzados a aproximarse entre sí debido al campo electrostático que se aplica entre la salida del capilar y la entrada al espectrómetro de masas ($\pm 2-5$ kV). En cierto momento, la repulsión de los iones se hace mayor que la tensión de la superficie que mantiene unidas las gotas en forma esférica, y las pequeñas gotas se rompen. Debido a fuerzas de repulsión coulombica aumenta la tensión superficial de las microgotas y éstas acaban "explotando" ("explosiones de Coulomb"), formándose así una serie de pequeñas gotas cargadas que seguirán sufriendo procesos de evaporación y explosión sucesivos hasta que finalmente se forman iones cargados desnudos que pasan a fase gaseosa con una o más cargas y son atraídos hacia la entrada del espectrómetro de masas como consecuencia del voltaje aplicado.



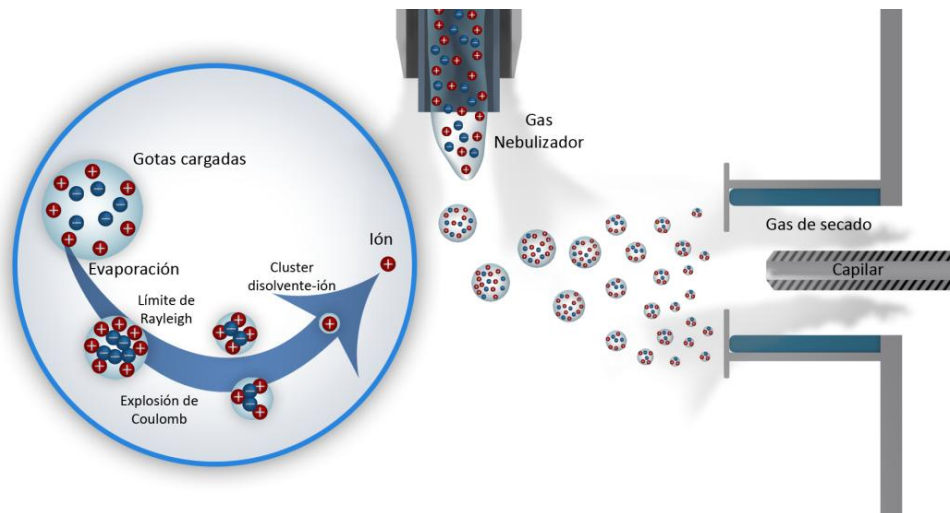


Figura 16. Esquema del proceso de formación del electrospray.

El hecho de que en el proceso de ionización se puedan formar iones mono- o multicargados, es una de las características que ha hecho que esta técnica haya revolucionado el campo de los acoplamientos entre técnicas separativas que trabajan en fase líquida y espectrometría de masas. Un incremento en el número de cargas en una molécula disminuirá el valor de su relación m/z , que es el parámetro utilizado para el análisis por un espectrómetro de masas. De este modo, mediante la detección de iones multicargados, se pueden detectar compuestos con pesos moleculares muy altos empleando analizadores de masas que trabajan con un intervalo limitado de valores m/z . Además, la ionización se puede llevar a cabo en el modo positivo o negativo. En el modo positivo, se podrán formar iones múltiplemente protonados $[M+nH]^{n+}$ donde n es el número de protones cargados positivamente en la molécula. Del mismo modo, es posible también la formación de aductos con iones sodio, litio, potasio, amonio, etc. En el modo negativo, pueden formarse iones $[M-nH]^{n-}$ debido a la desprotonación múltiple de las moléculas^{92,93}.

⁹² Simó, C., & Cifuentes, A., 2005. Electroforesis Capilar: detección mediante espectrometría de masas. In Universidad de Granada, ed. *Electroforesis capilar: aproximación según la técnica de detección*. pp. 409-438.

Después de la interfase, los iones son dirigidos hacia el **analizador de masas**, que permite la separación, detección y cuantificación de los analitos objeto de estudio con un grado de sensibilidad y selectividad muy elevado y proporcionando información sobre su masa molecular. Existen diversos tipos de analizadores, entre los que destacan el cuadrupolo (Q), el analizador de tiempo de vuelo (TOF) y el cuadrupolo-tiempo de vuelo (QTOF), dado que son los utilizados en el desarrollo experimental de esta memoria^{94,95}.

El analizador masas de **cuadrupolo** (Q) está formado por cuatro varillas paralelas equidistantes a un eje central imaginario (Figura 17). A las varillas opuestas se les aplica un potencial continuo de igual signo y opuesto al de las otras dos varillas. Adicionalmente, se aplica un potencial alterno asociado con una radiofrecuencia. En estas condiciones, los iones avanzarán a lo largo del analizador siguiendo trayectorias oscilantes, siendo repelidos y atraídos continuamente por las varillas. Para valores definidos de los voltajes continuo y alterno sólo atraviesan el cuadrupolo los iones con determinada relación m/z , los demás se desestabilizan y chocan contra las varillas. Por tanto, puede decirse que un cuadrupolo actúa como un filtro de masas. Al mantener constante la razón entre el potencial continuo y el alterno pero variando su intensidad, se consigue la salida secuencial de los iones con diferente m/z del cuadrupolo, generando así un espectro de masas^{96,97}.

⁹³ Smith, R. M., & Busch, K. L., 1999. Instrumentation. In J. Wiley & Sons, ed. *Understanding Mass Spectra. A basic approach*. New York, pp. 1-40.

⁹⁴ Marshall, A.G. & Hendrickson, C.L., 2008. High-resolution mass spectrometers. *Annual Review of Analytical Chemistry*, 1, pp.579–599.

⁹⁵ McLuckey, S. a & Wells, J.M., 2001. Mass analysis at the advent of the 21st century. *Chemical Reviews*, 101(2), pp.571–606.

⁹⁶ Finnigan, R.E., 1994. Quadrupole mass spectrometers. From development to commercialization. *Analytical Chemistry*, 66(19), pp.969–975.

⁹⁷ Miller, P.E. & Denton, M.B., 1986. The quadrupole mass filter: basic operating concepts. *Journal of Chemical Education*, 63(7), pp.617–622.



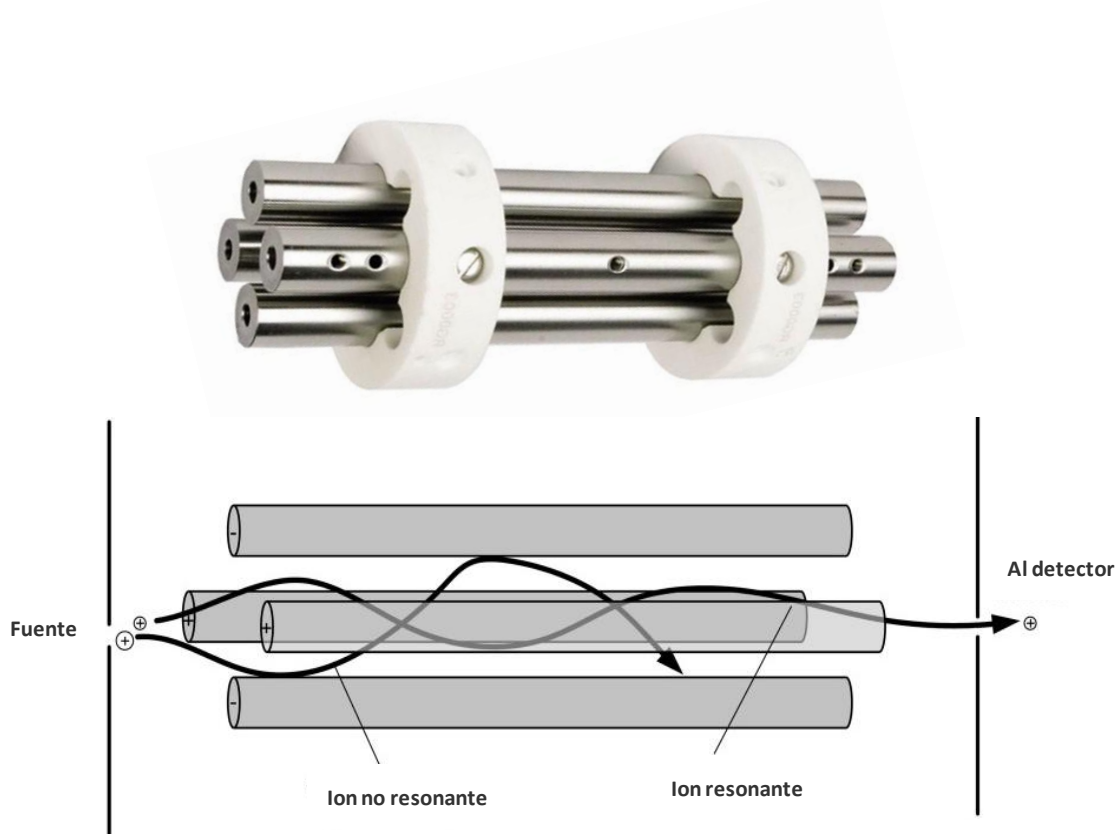


Figura 17. Esquema de un analizador de masas cuadrupolo (Q).

El analizador de masas de **tiempo de vuelo** (TOF) separa los iones según la distinta velocidad que adquieren en su interior en función de la relación m/z . En primer lugar, los iones son extraídos de la cámara de ionización y acelerados hacia el tubo de vuelo mediante un campo electrostático que les aporta una elevada energía cinética. Los iones de mayor m/z "volarán" a menor velocidad que los de menor m/z . La resolución entre los iones de diferente m/z será mejor cuanto mayor sea la longitud del tubo (habrá una mayor separación de los iones en el tiempo) y menor sea la dispersión en las energías de los iones formados en la fuente. En la Figura 18 se muestra el esquema del interior de un TOF^{98,99}.

⁹⁸ Verentchikov, a N., Ens, W. & Standing, K.G., 1994. Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. *Analytical Chemistry*, 66(1), pp.126–133.

⁹⁹ Rubinson K.A., Rubinson J.F., 2000. Espectrometría de masas. In Pearson Education, ed. *Análisis Instrumental*, pp. 522-577.

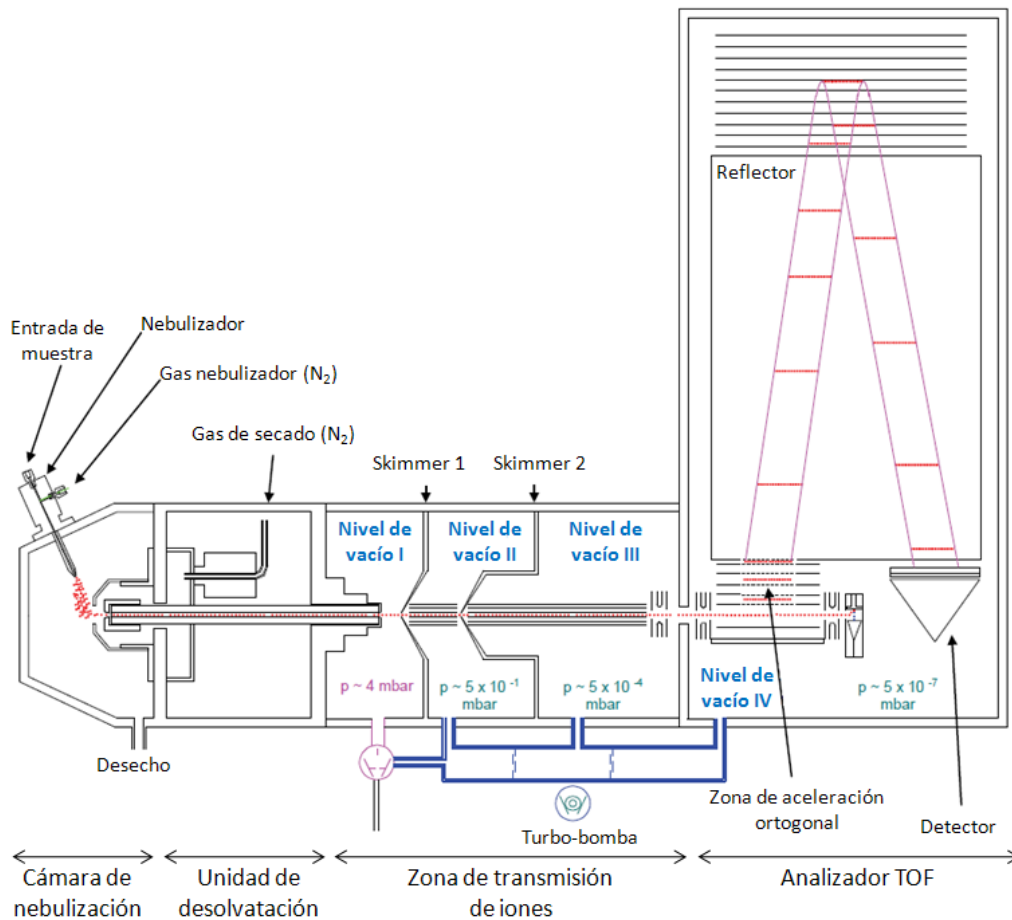


Figura 18. Esquema detallado de las partes de un espectrómetro de masas TOF.

Por orden, se encuentra la cámara de formación del spray, la unidad de desolvatación, la zona de transmisión o transferencia óptica, la zona de aceleración ortogonal y el analizador TOF propiamente dicho. Los iones formados en la cámara de nebulización atraviesan la unidad de desolvatación, que separa las zonas a presión atmosférica de la primera zona de alto vacío, y que consta de un calentador del gas de secado y un capilar de vidrio. Se llega a través de ella a la zona de transferencia óptica, que consta de tres módulos a alto vacío, separados entre sí por dos *skimmers*. El primero de los módulos (nivel de vacío I) está conectado a una bomba externa rotatoria, mientras que los otros módulos se conectan a una bomba turbo molecular que posee distintos niveles de vacío. Tanto en el nivel de vacío II como en el III, se

encuentra un hexapolo, encargado de transferir los iones hasta la zona de alto vacío. Al final de la zona de transmisión hay tres lentes que enfocan y dirigen dichos iones.

El analizador TOF consta a su vez de tres zonas. La primera de ellas es la zona de aceleración ortogonal donde se encuentran dos lentes que aceleran los iones hacia el tubo de vuelo, aplicando un campo eléctrico intermitente. Los iones volarán a lo largo del tubo de vuelo hasta alcanzar la segunda zona, el reflector del TOF (conocido como reflectrón), que permite corregir la dispersión en la energía cinética de los iones, aumentando de esta forma la resolución. El último componente del analizador TOF es el detector de impacto electrónico que consiste en una serie de placas a alto voltaje que convierten el impacto de los iones en señales eléctricas¹⁰⁰.

El analizador TOF es rápido y sensible, y su principal característica es que permite obtener valores de masa muy exactos, gracias a su elevada resolución. Las principales especificaciones del TOF son el amplio rango de masas en el que permite trabajar (50-3.000 Da), su resolución expresada como anchura de pico a mitad de su altura máxima (10.000-15.000 FWHM, del inglés *full width at half maximum*), y su exactitud (3 ppm y 5 ppm con calibración interna y externa, respectivamente).

El analizador de masas **cuadrupolo-tiempo de vuelo** (QTOF) está formado básicamente por las mismas partes que un analizador TOF, con la diferencia de que se introduce un cuadrupolo, donde se pueden seleccionar determinados iones para posteriormente ser fragmentados en una celda de colisión con ayuda de un gas de colisión (normalmente N₂). Los iones fragmentados se separan en el analizador TOF en función de su relación m/z de la misma forma que se ha

¹⁰⁰ micrOTOFcontrol 1.1. Operator Manual. (May 2005). Bruker Daltonics.



descrito en el apartado anterior. Además se sustituyen los *skimmers* de la zona de transmisión de iones por *funnels* (anillos concéntricos apilados en forma de embudo) que evitan la pérdida de iones durante la transmisión iónica, aumentando de esta forma la sensibilidad. En la Figura 19 se detallan las distintas partes de las que consta un analizador de masas QTOF¹⁰¹.

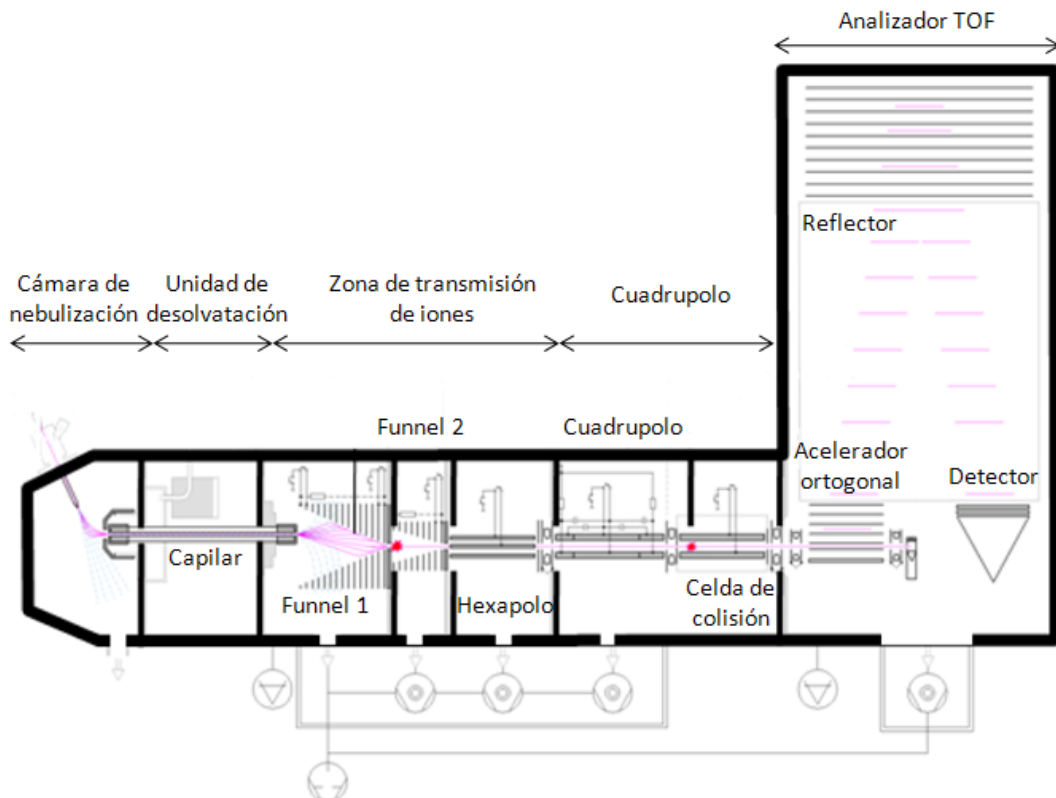


Figura 19. Esquema detallado de las partes de un espectrómetro de masas QTOF.

Las principales especificaciones de este analizador son un rango de masas de 50-20.000 m/z , una resolución de entre 17.500 y 20.000 FWHM, la posibilidad de llevar a cabo análisis de MS/MS, y una exactitud de 3 ppm con calibración

¹⁰¹ microTOFQ. User Manual. Version 1.1 (April 2007). Bruker Daltonics.



interna y 5 ppm con calibración externa, tanto de iones precursores como de fragmentos^{102,103}.

7. PROCESAMIENTO CONVENCIONAL DE LOS DATOS DE ESPECTROMETRÍA DE MASAS (MS).

Una vez analizadas las muestras vegetales mediante cromatografía líquida o de gases acoplada a espectrometría de masas, y debido a la complejidad de las muestras, se dispone de una gran cantidad de datos que hay que procesar. El procesamiento de los datos de espectrometría de masas se centra en un primer paso en la caracterización de los perfiles cromatográficos, ya sean polifenólicos o de compuestos volátiles. Normalmente, se realiza primero una caracterización cualitativa, que consiste en identificar los compuestos de las muestras analizadas y obtener una lista de compuestos identificados con los distintos parámetros que describen a cada uno. Posteriormente, puede llevarse a cabo la caracterización cuantitativa o cuantificación de los compuestos identificados, junto con los estudios estadísticos pertinentes.

El procesamiento de los datos se realiza convencionalmente con el software propio que proporciona la casa comercial del equipo utilizado en el análisis. A continuación, se detalla el método que se utiliza convencionalmente para la caracterización de los datos de MS, tanto para GC como para LC, destacando la información procedente del equipamiento utilizado en el desarrollo experimental de la tesis.

De los análisis de **cromatografía de gases** acoplada a **espectrometría de masas** (GC-MS), se obtienen los perfiles cromatográficos que proporcionan

¹⁰² Bristow, T. *et al.*, 2008. Performance optimisation of a new-generation orthogonal-acceleration quadrupole-time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry*, 22, pp.1213–1222.

¹⁰³ Chernushevich, I. V, Loboda, a V & Thomson, B. a, 2001. An introduction to quadrupole-time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, 36(8), pp.849–865.



información sobre el tiempo de retención de cada compuesto y su espectro de masas, consistente en las características de su patrón de fragmentación. Cuando la interfase y el analizador de masas utilizados son de impacto electrónico y de cuadrupolo, respectivamente, la fragmentación obtenida es de gran ayuda en la identificación, sin embargo, la información de masa proporcionada no es exacta. Esto no supone un gran obstáculo en la caracterización puesto que, en GC-MS contamos con una amplia variedad de librerías tales como las bases de datos Wiley y NIST MS^{74,99}.

La caracterización a partir de datos de **cromatografía líquida** acoplada a **espectrometría de masas** (LC-MS), se basa en la determinación de masas exactas y el uso del patrón de distribución isotópica o TIP (*true isotopic pattern*), cuando el analizador de masas utilizado es de tiempo de vuelo (TOF). A partir de ambos datos, el software utiliza un algoritmo que proporciona una lista de posibles fórmulas moleculares para cada compuesto, así como también una sofisticada comparación entre el patrón isotópico teórico y el experimental para aumentar la confianza en la fórmula molecular sugerida¹⁰⁰.

Cuando la interfase del TOF es de ionización por electrospray, el espectro de masas resultante consta del pico del ión molecular y sólo alguno o incluso ningún otro pico. A pesar de constituir un sistema de ionización blando, se detectan a menudo ciertas fragmentaciones producidas en la fuente de ionización, principalmente, cuando se trata de compuestos lábiles y/o de alto peso molecular. Dicha fragmentación se puede definir como disociación inducida por colisión en la fuente o ISCID (in-source collision-induced dissociation)¹⁰⁴.

¹⁰⁴ El-Aneed, A., Cohen, A. & Banoub, J., 2009. Mass spectrometry, review of the basics: electrospray, maldi, and commonly used mass analyzers. *Applied Spectroscopy Reviews*, 44(3), pp.210–230.



El analizador de masas QTOF ofrece una ventaja adicional con respecto al TOF. En este caso, el cuadrupolo actúa como celda de colisión, de manera que la caracterización no sólo se basa en la masa exacta y la distribución isotópica, sino que además aporta información estructural procedente del patrón de fragmentación de los compuestos se obtiene la masa exacta de cada fragmento¹⁰⁵. Los análisis de MS/MS obtenidos con el QTOF ofrecen ventajas con respecto a analizadores de masas tales como la trampa de iones o el cuadrupolo. Con el QTOF, la fragmentación es selectiva, garantizando que los fragmentos obtenidos proceden de una molécula exclusivamente. Además, aporta la masa exacta de los mismos.

En el proceso de caracterización de muestras analizadas por LC-MS, el siguiente paso es la búsqueda de la fórmula molecular en bases de datos de compuestos. Existen múltiples bases de datos de compuestos químicos entre las que destacan *SciFinder Scholar*, *Kegg Ligand*, *PubChem*, *ChemIDplus Advanced* o *ChemSpider*. Una vez que se tienen identificadas las múltiples moléculas que presenten la fórmula molecular asignada, en el caso que corresponda, se estudian los fragmentos del espectro de masas. Para ello, se tienen en cuenta las reglas de fragmentación y/o la fragmentación previamente descrita en bibliografía para las moléculas candidatas. También pueden consultarse las bases de datos de espectros de masas como *MassBank* o *METLIN* que almacenan espectros de masas de distintas sustancias. Otra herramienta web que puede ser útil es la denominada *MetFrag*, que calcula fragmentaciones teóricas de las moléculas encontradas en bases de datos de compuestos químicos y las compara con la fragmentación que el usuario introduzca^{106,107}.

¹⁰⁵ Aebersold, R. & Mann, M., 2003. Mass spectrometry-based proteomics. *Nature*, 422(6928), pp.198–207.

¹⁰⁶ Carroll, A.J., 2012. Online metabolomics databases and pipelines. In U. Roessner, ed. *Metabolomics*. InTech, pp. 47–72.



8. NUEVAS HERRAMIENTAS DE PROCESAMIENTO DE DATOS DE ESPECTROMETRÍA DE MASAS (MS).

La tarea más importante del análisis químico de una muestra consiste en la determinación de la composición elemental de todos los compuestos conocidos o desconocidos presentes en la muestra. Cuando el objetivo es identificar compuestos desconocidos, la tarea se complica, ya que requiere la determinación de masas exactas con alta precisión. Además, en los últimos años el descubrimiento del importante papel que pueden jugar las moléculas minoritarias en la química de los sistemas biológicos, ha conducido al crecimiento del campo de la metabolómica y ha generado la necesidad de análisis que sean capaces de identificar el mayor número de compuestos posible en una muestra^{108,109}.

La aparición de la técnica analítica consistente en el acoplamiento de cromatografía acoplada a espectrometría de masas (LC/GC-MS), como técnica para la identificación y cuantificación ha supuesto un gran avance gracias a la gran sensibilidad y reproducibilidad que proporcionan^{110,111}.

El desarrollo de la técnica LC/GC-MS, a partir de los años 90, se ha visto acompañado por el desarrollo de muchos algoritmos para procesar los datos. Los algoritmos de procesamiento de los datos LC/GC-MS se introducen en

¹⁰⁷ Thurman, E.M., Ferrer, I. & Fernández-Alba, A.R., 2005. Matching unknown empirical formulas to chemical structure using LC/MS TOF accurate mass and database searching: example of unknown pesticides on tomato skins. *Journal of Chromatography A*, 1067(1-2), pp.127–134.

¹⁰⁸ Nicholson, J.K., Lindon, J.C. & Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11), pp.1181–1189.

¹⁰⁹ Fiehn, O., 2002. Metabolomics-the link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1-2), pp.155–171.

¹¹⁰ Bruins, A.P., Covey, T.R. & Henion, J.D., 1987. Ion spray interface for combined liquid chromatography/atmospheric pressure ionization mass spectrometry. *Analytical Chemistry*, 59(22), pp.2642–2646.

¹¹¹ Arpino P. Combined liquid chromatography mass spectrometry. Part III. Applications of thermospray. *Mass Spectrom. Rev.*, 1992, 11, 3-40.



paquetes de programas de software. Las casas comerciales de los equipos de MS proporcionan su propio software, pero ninguno presenta una solución general al problema de la determinación de picos cromatográficos. Todos los métodos de procesamiento dependen del tipo de instrumento específico y de la naturaleza de la muestra a tratar. Además, requiere de un análisis pormenorizado de cada uno de los picos del perfil cromatográfico de forma manual por el usuario. El principal inconveniente de este procedimiento es que no detecta los picos que coeluyen, situación que ocurre frecuentemente ya que la técnica cromatográfica no es capaz de separar todos los compuestos que componen una muestra compleja. Además, con esta técnica se invierte mucho tiempo computacional en el análisis de cada muestra, lo que dificulta el procesamiento de colecciones de muchas muestras.

Para el análisis de muestras complejas de forma automática, distintos centros de investigación a nivel mundial están desarrollando software libre adaptado a las especificaciones concretas de cada estudio¹¹²⁻¹¹⁵. Este software libre está disponible en Internet y trabaja bajo distintas plataformas matemáticas y con distintos formatos de los ficheros a analizar. En los trabajos de investigación de Goodacre *et al.* (2007) y Want *et al.* (2011) se hace una revisión detallada de todos los paquetes de software disponibles en el mercado y se introduce la terminología para referenciar cada proceso que realizan los algoritmos^{116,117}.

¹¹² Sturm, M. *et al.*, 2008. OpenMS-An open-source software framework for mass spectrometry. *BMC bioinformatics*, 11, pp.1–11.

¹¹³ Yu, T. *et al.*, 2009. apLCMS-adaptive processing of high-resolution LC/MS data. *Bioinformatics*, 25(15), pp.1930–1936.

¹¹⁴ Zhang, J. *et al.*, 2009. Review of peak detection algorithms in liquid-chromatography-mass spectrometry. , 10, pp.388–401.

¹¹⁵ Tautenhahn, R., Böttcher, C. & Neumann, S., 2008. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 16, pp.1–16.

¹¹⁶ Goodacre, R. *et al.*, 2007. Proposed minimum reporting standards for data analysis in metabolomics. *Metabolomics*, 3, pp.231–241.

¹¹⁷ Want, E. & Masson, P., 2011. Processing and analysis of GC/LC-MS-based metabolomics data. In S. Protocols, ed. *Metabolic Profiling, Methods in Molecular Biology*. pp. 277–298.



Son muchas las herramientas de procesamiento automáticos de los datos LC-MS disponibles, y se diferencian en el tipo de muestra a analizar (proteínas, péptidos, plantas, metabolitos,..), en la forma de abordar el problema (detección de picos cromatográficos mediante diferente modelos de gaussianas, analizando la forma de los picos,..), en que se centran en un solo paso del análisis (reducción de ruido, cuantificación, alineamiento de distintas muestras,..) o en una combinación de pocos pasos.

La gran mayoría de las herramientas de procesamiento automático están basadas, independientemente de la naturaleza de la muestra a analizar, en el análisis de picos cromatográficos o lo que se conoce como *Peak Picking* o *Peak Detection*, y proponen diferentes soluciones a problemas del ruido, de la coelución, del desalineamiento, etc^{115,118-123}.

La tendencia actual va encaminada a desarrollar herramientas que estudien los datos LC-MS en su globalidad, visualizando y tratando los datos en tres dimensiones, es decir, considerando la intensidad de los picos en función de la masa y el tiempo de retención de cada compuesto. Hay varias referencias en bibliografía que han comenzado a tratar el problema desde este punto de

¹¹⁸ Morris, J.S. *et al.*, 2005. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, 21(9), pp.1764–1775.

¹¹⁹ Du, P., Kibbe, W. a. & Lin, S.M., 2006. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17), pp.2059–2065.

¹²⁰ Lange, E. *et al.*, 2006. High-accuracy peak picking of proteomics data using wavelet techniques. *Pacific Symposium on Biocomputing*, 11, pp.243–254.

¹²¹ Yan, G. *et al.*, 2010. Rapid and global detection and characterization of *aconitum* alkaloids in Yin Chen Si Ni Tang, a traditional Chinese medical formula, by ultra performance liquid chromatography-high resolution mass spectrometry and automated data analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 53(3), pp.421–431.

¹²² Fredriksson, M.J. *et al.*, 2011. Combined use of algorithms for peak picking, peak tracking and retention modelling to optimize the chromatographic conditions for liquid chromatography-mass spectrometry analysis of fluciclonolone acetonide and its degradation products. *Analytica Chimica Acta*, 704, pp.180–188.

¹²³ Krishnan, S. *et al.*, 2012. Instrument and process independent binning and baseline correction methods for liquid chromatography-high resolution-mass spectrometry deconvolution. *Analytica Chimica Acta*, 740, pp.12–19.



vista¹²⁴⁻¹²⁶. En la bibliografía se suele presentar la herramienta desarrollada para el procesamiento de los datos LC/GC-MS y los resultados se suelen comparar con los obtenidos, utilizando el software propio del equipo proporcionado por la casa comercial, lo que denominamos método convencional. Se presentan los resultados y se analizan las mejoras con respecto al método convencional.

En la presente tesis doctoral se han desarrollado nuevas herramientas automáticas de procesamiento de los datos LC-MS para el análisis de compuestos bioactivos en extractos vegetales. Están específicamente diseñadas para procesar los datos procedentes de los equipos utilizados en el desarrollo experimental de esta tesis (Bruker Daltonik), pero son fácilmente adaptables al procesamiento de datos de otros equipos, mediante una simple transformación de los formatos de los ficheros de salida de los equipos.

El objetivo último de cualquier estudio de análisis químico es, a partir de los datos proporcionados por LC-MS, obtener una lista de masas exactas y tiempos de retención, áreas o intensidades iónicas y patrones espectrométricos para cada uno de los compuestos que formen la muestra, solucionando los problemas de coelución, ruido, desalineamiento, etc. En esta tesis, se abordan todos estos problemas desde un punto de vista global de los datos, considerando tres dimensiones (tiempo de retención, masa, intensidad) de forma automática.

Una de las ventajas del método propuesto reside en el preprocesamiento de los datos de salida del LC-MS, que consiste en la reducción del tamaño de los

¹²⁴ Stolt, R. *et al.*, 2006. Second-order peak detection for multicomponent high-resolution LC/MS data. *Analytical Chemistry*, 78(4), pp.975–983.

¹²⁵ Katajamaa, M. & Oresic, M., 2007. Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, 1158(1-2), pp.318–328.

¹²⁶ Stee, L.L.P. Van & Brinkman, U.A.T., 2011. Peak clustering in two-dimensional gas chromatography with mass spectrometric detection based on theoretical calculation of two-dimensional peak shapes: The 2DAid approach. *Journal of Chromatography A*, 1218(43), pp.7878–7885.



ficheros de las muestras (mediante el filtrado adecuado para la eliminación del ruido) y en la calibración de masa. La reducción del tamaño de los ficheros hace que su procesamiento sea mucho más eficaz, en la medida en que facilita cualquier cálculo matemático posterior, haciéndolo mucho más rápido. El tiempo requerido se reduce considerablemente, suponiendo una ventaja de vital importancia en los estudios que abarcan un gran número de muestras.

Una vez realizado el preprocesamiento de los datos, se está en las mejores condiciones para analizar los datos y conseguir el principal objetivo que es obtener un listado de compuestos presentes, junto con sus características principales (área, masa exacta, tiempo de retención, desviación estándar de masa y tiempo...) de cada una de las muestras. El método de procesamiento propuesto se basa en una idea simple y efectiva, en la idea de "divide y vencerás". Consiste en dividir sucesivamente el espectro de masas en unidades pequeñas y analizarlas separadamente, hasta llegar a las unidades más pequeñas posibles, las cuales conformarán las distintas especies iónicas. Esto se realiza en diferentes fases, y con diferentes procedimientos. El listado de compuestos generado por la herramienta propuesta tiene forma de tabla o matriz de datos, de manera que proporciona directamente toda la información necesaria de los compuestos detectados, así como el valor de área calculada de manera automática para cada compuesto en cada muestra.

Este nuevo procesamiento de los datos de MS que se propone permite una visión más amplia de los análisis de las muestras. Es especialmente útil en la caracterización de compuestos en muestras vegetales, donde debido a su complejidad, la tendencia del investigador es identificar los picos mayoritarios del perfil cromatográfico, con la consiguiente posibilidad de estar obviando picos de menor intensidad que, por coelución, se encuentren solapados debajo de picos de mayor intensidad.



Pero estas nuevas herramientas de procesamiento no sólo pueden aplicarse a la caracterización de compuestos bioactivos en muestras vegetales, sino que dado el gran tamaño de los ficheros procedentes de estudios de proteómica, se considera su uso en este campo realmente necesario y ventajoso. Puede aplicarse también al estudio de muestras biológicas, en los que cualquier compuesto, que por intensidad pueda confundirse con el ruido, puede ser decisivo. Concretamente, la confianza de no perder información es muy útil en la realización de análisis diana o dirigidos (lo que se conoce como *target* análisis) ya que se centran exclusivamente en un metabolito concreto que resulte de interés. En general, el uso de la herramienta propuesta en estudios de metabolómica es muy recomendable dado que el número de muestras utilizadas en estos estudios es muy elevado y se requiere la comparación entre las muestras, lo cual supone un consumo muy elevado de tiempo computacional y de personal con los métodos convencionales de procesamiento de datos.

Otra aplicación interesante que ofrece la herramienta de preprocesamiento de datos de MS propuesta es que permite el desarrollo directo de una amplia gama de cálculos estadísticos, tales como análisis de componentes principales (PCA), análisis discriminante lineal (LDA), etc. Cuando estos cálculos se realizan manejando los datos preprocesados, se obtienen los resultados en cuestión de pocos minutos, lo cual es una gran ventaja respecto al método convencional.

El nuevo método de preprocesamiento de los datos de MS destaca por ser automático, rápido, versátil (sirve tanto para datos de GC como para datos de LC), está muy controlado por el usuario y además se puede adaptar a cada estudio específico con sus particularidades concretas.



La aplicación de este nuevo método a los estudios realizados en nuestro centro de investigación, que previamente se habían estudiado con el método convencional, ha proporcionado muy buenos resultados.

9. IDENTIFICACIÓN DE COMPUESTOS BIOACTIVOS.

El descubrimiento de compuestos biológicamente activos parte de la selección de plantas, la obtención de colecciones interesantes de extractos a partir de éstas y la valoración de la bioactividad de los extractos. Además, la probabilidad de identificar compuestos bioactivos aumenta conforme mayor sea el número de muestras objeto de estudio.

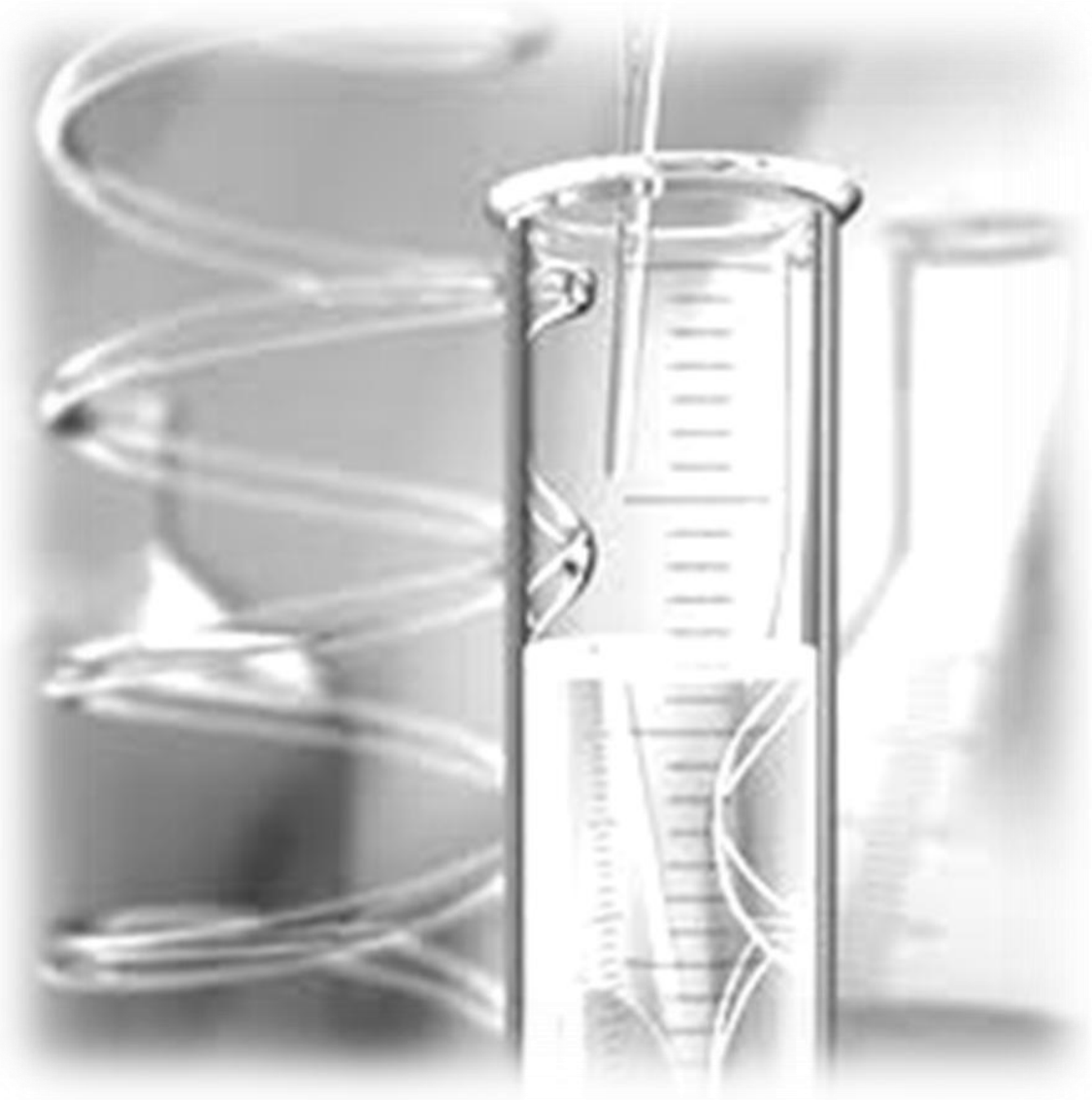
Cuando en el cribado de una gran variedad de muestras se encuentra una muestra de actividad biológica destacable, se hace necesario conocer qué compuesto/s es/son el/los responsable/s de la misma. Para ello, convencionalmente se recurre a un proceso de fraccionamiento biodirigido del extracto, consistente en fraccionar el extracto, coleccionar las distintas fracciones y valorar la bioactividad de cada una de ellas. El paso limitante de este proceso es la etapa de fraccionamiento, en la que el extracto es fraccionado y refraccionado hasta obtener una fracción bioactiva y pura. También es importante tener en cuenta el efecto adicional que suele producirse en los extractos sometidos a fraccionamiento ya que dicho proceso frecuentemente conlleva a una reducción o pérdida de la actividad biológica, dificultando así la identificación de la fracción bioactiva. Por lo tanto, sería deseable determinar los principios bioactivos del extracto sin la necesidad de fraccionarlo previamente.

La eficacia de las técnicas analíticas disponibles actualmente para la identificación de compuestos bioactivos depende fuertemente de la complejidad de las muestras y del esfuerzo implicado en el fraccionamiento y la



evaluación de la bioactividad. Con el objetivo de solventar o acotar los problemas asociados a este procedimiento convencional descrito anteriormente, se propone una herramienta estadística de correlación entre los datos de espectrometría de masas y los datos de bioactividad a partir de una amplia colección de muestras. Esta herramienta proporciona de forma automática una lista de compuestos candidatos o potencialmente responsables de dicha propiedad bioactiva, de manera que únicamente entrarían en el proceso de fraccionamiento aquellos que presentan una correlación más favorable entre los datos de espectrometría de masas y los datos de bioactividad, sin la necesidad de chequear la bioactividad de cada fracción.





METODOLOGÍA

La identificación del/los compuesto/s potencialmente responsable/s de la bioactividad se llevó a cabo mediante la utilización de una gran variedad de fuentes vegetales, diferentes sistemas de extracción, diferentes plataformas analíticas y diversos ensayos de bioactividad. Para ello, se siguió una metodología general durante el desarrollo experimental de la presente tesis, aplicando dicha metodología en los tres bloques que conforman la sección de resultados. La aplicación de la metodología general a los diferentes estudios realizados se presenta esquemáticamente en la Figura 20.

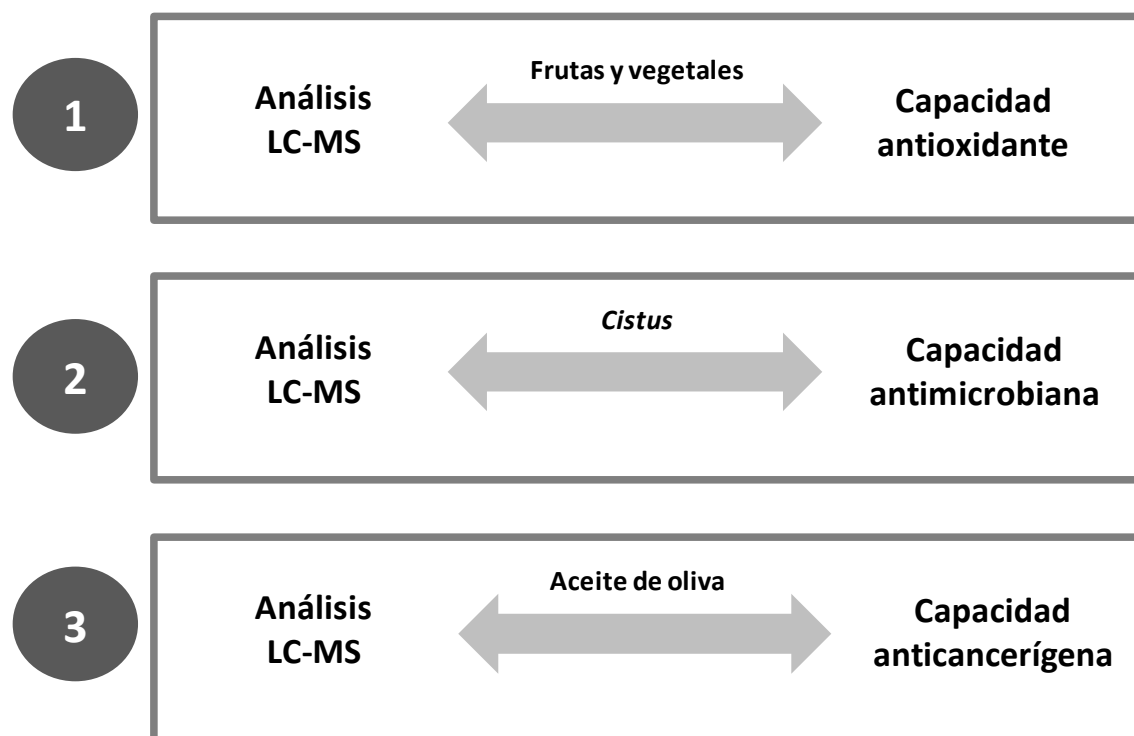


Figura 20. Esquema de los estudios realizados en la tesis.

En el primer estudio, se estudiaron frutas y vegetales, relacionando los datos del análisis LC-MS con los datos de capacidad antioxidante. En el segundo, se evaluó la capacidad antimicrobiana de una amplia colección de extractos polifenólicos de *Cistus* y se correlacionaron dichos datos con los correspondientes datos procedentes del análisis químico. En el tercer y último estudio, la propiedad

biológica estudiada fue la capacidad anticancerígena para correlacionarla con la composición química de diferentes extractos polifenólicos de aceite de oliva.

En la Figura 21 se presenta el esquema de la metodología general aplicada en la tesis, consistente en seis etapas. Se basó en una primera etapa de búsqueda y selección de fuentes vegetales interesantes para distintas aplicaciones (A), seguida de una etapa de extracción de los compuestos de interés, potencialmente bioactivos (B). A continuación, se analizó la composición de las matrices vegetales seleccionadas (C) y se evaluaron las propiedades bioactivas de las mismas (D). Los datos del análisis químico se procesaron mediante distintos métodos (E), y una vez procesados, dichos datos se correlacionaron con los datos obtenidos en los ensayos de bioactividad (F). Entre los resultados obtenidos, destaca la mejora en cuanto a la caracterización de compuestos y la obtención de compuestos candidatos a ser responsables de la bioactividad.

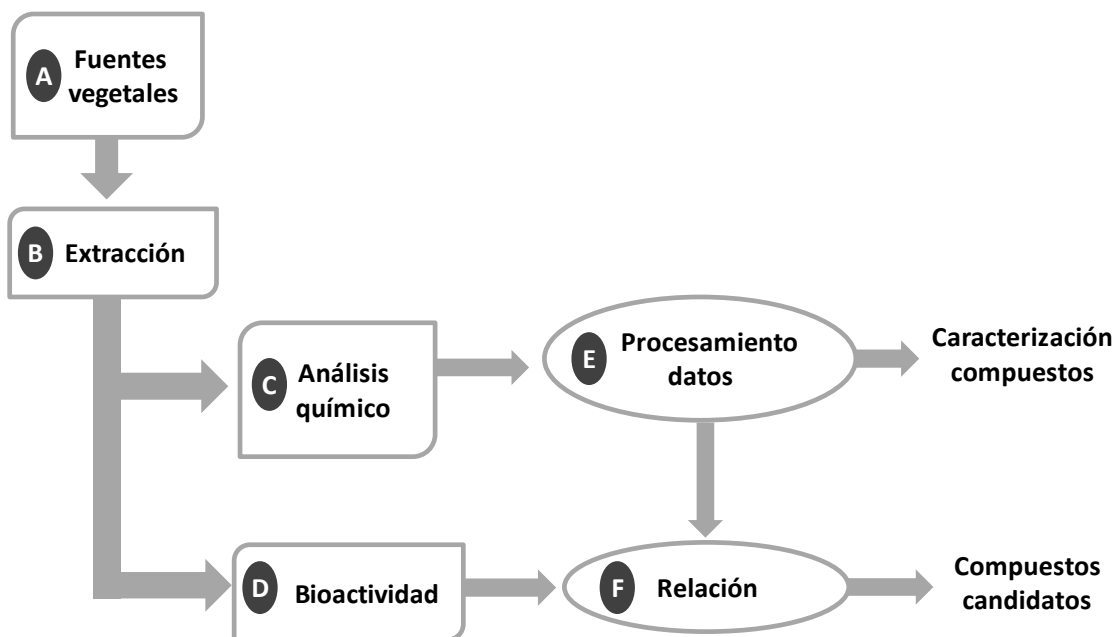


Figura 21. Esquema general de la metodología llevada a cabo.

De este modo, los datos a utilizar en los estudios de correlación para la identificación de candidatos abarcaron diferentes propiedades bioactivas y distintos tipos de compuestos que, dentro de la familia de los compuestos fenólicos, son característicos de cada fuente objeto de estudio. Asimismo, los datos de composición química procedieron de análisis de cromatografía líquida acoplada a espectrometría de masas (LC-MS), aunque con ciertas diferencias de equipamiento dependiendo de cada estudio.

A continuación, se describe de manera detallada la metodología desarrollada, siguiendo el orden de las etapas (apartados A – F). Cada apartado se presenta dividido a su vez en tres subapartados, dedicados a la descripción de las especificaciones metodológicas de cada uno de los estudios realizados.

A) SELECCIÓN DE MUESTRAS VEGETALES.

A.1) Selección de frutas y vegetales.

Para el estudio de la capacidad antioxidante, se seleccionaron veinticuatro especies distintas de frutas y vegetales. A continuación, se detallan las especies seleccionadas con su correspondiente nombre científico.

- Las especies de frutas seleccionadas fueron: aguacate (*Persea americana* Mill.), caqui (*Diospyros kaki*), chirimoya (*Annona cherimola* Mill.), granada (*Punica granatum* L.), limón (*Citrus limon* (L.) Burm. f.), mango (*Mangifera indica* L.), melón (*Cucumis melo* L.), membrillo (*Cydonia oblonga* Mill.), níspero (*Eriobotrya japonica* (Thunb.) Lindl.), sandía (*Citrullus lanatus* (Thunb.) Mansf.) y uva (*Vitis vinifera* L.).
- Las especies de vegetales seleccionados fueron: ajo (*Allium sativum* L.), alcachofa (*Cynara scolymus* L.), berenjena (*Solanum melongena* L.), calabacín (*Cucurbita pepo* L.), cebolla (*Allium cepa* L.), espárrago



(*Asparagus officinalis* L.), haba (*Vicia faba* L.), judía verde (*Phaseolus vulgaris* L.), lechuga (*Lactuca sativa* L.), patata (*Solanum tuberosum* L.), pimiento (*Capsicum annuum* L.), tomate (*Solanum lycopersicum* L.), y zanahoria (*Daucus carota* L.).

De cada especie, se seleccionaron las variedades más representativas del sureste de España, incluyendo entre una y tres variedades por especie. De esta forma, se obtuvo un total de cuarenta y cuatro variedades.

El muestreo se realizó a lo largo del año 2011. Se recolectó alrededor de 1.5 Kg de cada variedad, obteniendo muestras a distintos tiempos del periodo de cosecha de cada cultivo (entre dos y cinco tiempos). En la Figura 22 se muestran de manera esquemática los nombres de todas las variedades de frutas y vegetales que conforman el estudio, así como la temporada en la que están disponibles en el mercado, según los datos del Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA).



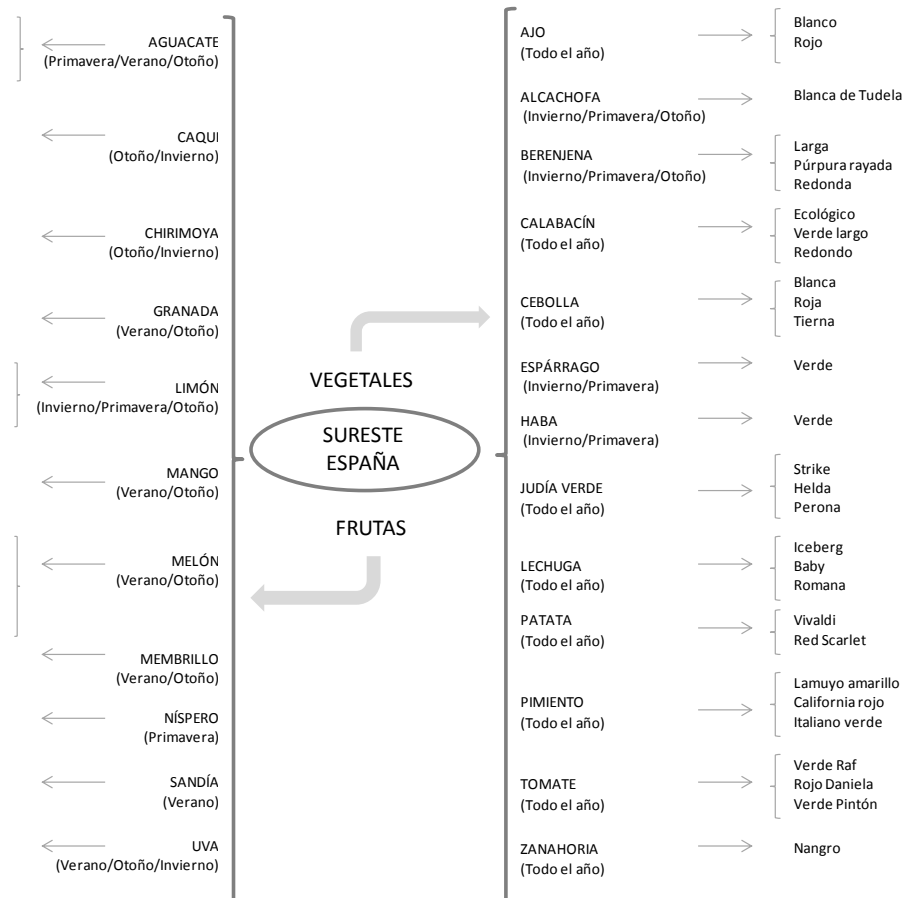


Figura 22. Esquema de todas las variedades de frutas y vegetales que conforman el estudio de capacidad antioxidante y la temporalidad de las mismas.

El lugar de recogida de las muestras es diferente en función de las variedades de frutas y vegetales. La gran mayoría de los vegetales proceden de Andalucía, predominantemente Almería, incluyendo distintas zonas o comarcas. Los tomates son de La Cañada (Almería) y los pimientos de San Agustín (Almería), a excepción de la variedad de pimiento verde que se adquirió en una tienda ecológica de Almería denominada "Semilla Verde". Las muestras de calabacín se obtuvieron de Las Norias (Almería), a excepción de la variedad ecológica que se compró en la tienda anteriormente citada. Las muestras de berenjena y judía son de San Agustín, excepto la variedad redonda de berenjena y la judía "Helda", que fueron adquiridas en Las Norias. Otras muestras de vegetales también procedentes de Almería son la zanahoria, el ajo, el espárrago y el haba,

concretamente esta última de Aguadulce. Las muestras de lechuga son de Níjar (Almería), a excepción de la variedad "Baby", obtenida en Murcia. Las variedades incluidas de alcachofa y patata son de Granada. Las muestras de cebolla son de procedencia dispar, de Valencia la variedad blanca, de Murcia la variedad roja y de Almería la cebolla tierna o cebolleta.

En el caso de las frutas seleccionadas, la gran mayoría son de Granada, como el aguacate, el níspero, la granada y el membrillo (este último concretamente de la Alpujarra). Las muestras de limón, sandía, melón y uva son de Almería, concretamente de El Ejido todas las variedades de melón y sandía, y de Berja la uva roja. Por último, las muestras de mango fueron obtenidas en Málaga, las de caqui en Valencia y las de chirimoya en Murcia.

De todas las muestras, se seleccionaron las muestras de pimiento para un estudio pormenorizado de sus constituyentes fenólicos. Se trata de tres variedades de pimiento: "dulce Italiano", "Lamuyo" y "California wonder". La primera es una variedad ecológica característica por su forma alargada y su color verde. Las otras dos variedades tienen la forma típica de pimiento más redondeada con colores amarillo y rojo, respectivamente. Se cultivaron bajo las mismas condiciones agronómicas, siendo recolectadas entre septiembre y octubre de 2011.

A.2) Selección de especies del género *Cistus*.

La selección de las muestras de *Cistus* se realizó en colaboración con el Instituto de Biología Molecular y Celular (IBMC) de la Universidad Miguel Hernández de Elche (Alicante, España), así como la empresa QUÍMICAS DEL VINALOPÓ, S.L. en Monóvar (Alicante, España). La selección vegetal se basó en la abundancia de las especies, su adaptación al clima y a las condiciones del suelo de las regiones semiáridas del sureste de España, así como en la facilidad de los procedimientos de cultivo. Se seleccionaron cuatro especies diferentes: *Cistus ladanifer* L., *Cistus*



albidus L., *Cistus salviifolius* L. y *Cistus clusii* Dunal, las cuales formaron parte tanto del estudio de la fracción polar como de la fracción apolar. En ambos estudios, la extracción se realizó con la parte aérea de la planta.

Las plantas se obtuvieron de diferentes áreas de España. Las muestras de *C. ladanifer* se obtuvieron de un área montañosa cercana a Puertollano, provincia de Ciudad Real, cerca de Andalucía. Las muestras de *C. albidus*, *C. salviifolius* y *C. clusii* se recogieron de un área semiárida cercana a Monóvar, provincia de Alicante.

Con el fin de aumentar el número de muestras incluidas en el estudio de correlación entre datos de análisis LC-MS y datos de bioactividad, se añadieron dos especies al estudio: *C. psilosepalus* y *C. populifolius*.

El material vegetal concerniente al estudio del perfil polifenólico y la actividad antimicrobiana de todas las muestras fue recolectado entre junio de 2010 y noviembre de 2011. Para el estudio de los compuestos volátiles, las muestras se recolectaron en junio de 2013.

A.3) Selección de muestras de aceite de oliva virgen extra.

En el proceso de selección de muestra para este estudio, se consideraron veintidós aceites de oliva virgen extra (EVOO)^{29,127}. Por un lado, se seleccionaron catorce aceites monovarietales de alto contenido polifenólico de procedencia controlada y garantía de homogeneidad, contactando con diferentes compañías comercializadoras de aceite, almazaras y denominaciones de origen. Estos aceites fueron de diferentes variedades y distintas zonas geográficas de España: dos aceites de oliva Hojiblanca producidos en Málaga (1) y Sevilla (9), siete aceites de oliva Picual producidos en Málaga (2), Jaén (4, 10 y 11), Granada (5 y

¹²⁷ García-Villalba, R. *et al.*, 2010. Characterization and quantification of phenolic compounds of extra-virgin olive oils with anticancer properties by a rapid and resolute LC-ESI-TOF MS method. *Journal of Pharmaceutical and Biomedical Analysis*, 51(2), pp.416–429.



6), y Córdoba (7), un Cornezuelo (3), un Manzanilla (8), y tres aceites de oliva Arbequina (12, 13 y 14). Las aceitunas se procesaron en continuo en plantas industriales y se almacenaron en botellas sin espacio de cabeza a temperatura ambiente y en ausencia de luz. Todos los aceites se produjeron en el mismo año (2008), se congelaron a -20°C para los análisis posteriores. Por otro lado, las ocho muestras restantes de la colección fueron aceites de oliva comerciales (codificados del 15 al 22), abarcando las variedades Picual, Hojiblanca y Arbequina.

La colección completa de aceites de oliva virgen extra se dividió al azar en dos grupos para la posterior aplicación de la herramienta de correlación: un grupo de "entrenamiento", utilizados para la identificación de compuestos candidatos y la estimación del modelo predictivo, y otro grupo de muestras de "evaluación", utilizados para evaluar el modelo de actividad. Toda la información referente a las variedades de los aceites de oliva, las zonas de los mismos y las muestras asignadas a cada grupo se detalla en la Tabla 1.



Tabla 1. Extractos polifenólicos de aceite de oliva virgen extra de diferentes variedades de aceituna, diferentes aéreas geográficas de España.

Extracto	Variedad	Zona geográfica	Grupo
1	Hojiblanca	Málaga	Entrenamiento
2	Picual	Málaga	Entrenamiento
3	Cornezuelo	Badajoz	Entrenamiento
4	Picual	Jaén	Entrenamiento
5	Picual	Granada	Evaluación
6	Picual	Granada	Entrenamiento
7	Picual	Córdoba	Evaluación
8	Manzanilla	Sevilla	Evaluación
9	Hojiblanca	Sevilla	Evaluación
10	Picual	Jaén	Entrenamiento
11	Picual	Jaén	Evaluación
12	Arbequina	Reus	Entrenamiento
13	Arbequina	Sevilla	Evaluación
14	Arbequina	Sevilla	Entrenamiento
15	Picual	Comercial	Entrenamiento
16	Picual	Comercial	Entrenamiento
17	Arbequina	Comercial	Entrenamiento
18	Arbequina	Comercial	Entrenamiento
19	Hojiblanca	Comercial	Entrenamiento
20	Hojiblanca	Comercial	Evaluación
21	Hojiblanca	Comercial	Entrenamiento
22	Picual	Comercial	Evaluación

"Entrenamiento", extractos para identificación de candidatos y estimación del modelo predictivo.

"Evaluación", extractos para evaluar el modelo de actividad.



B) EXTRACCIÓN DE COMPUESTOS BIOACTIVOS.

B.1) Extracción de compuestos antioxidantes de frutas y vegetales.

Para la preparación de extractos de frutas y vegetales, para los cuales se evaluará la capacidad antioxidante, primeramente, se separó la porción comestible de cada muestra fresca, desechando las semillas y la piel (en los casos en los que no forman parte de la porción comestible). Después, las muestras se trituraron, homogeneizaron y almacenaron a -20°C para llevarlas posteriormente al equipo de liofilización (Christ Alpha 1-2 LD Freeze dryer, Shropshire, Reino Unido), el cual se acondicionó previamente a -50°C y 1 mbar durante un hora. Antes de la liofilización, se determinó el porcentaje de humedad de cada muestra.

Para cada muestra, se pesaron 0.5 g en un tubo de centrífuga y se añadieron 15 mL de metanol/agua 80:20 (v/v). La mezcla se sometió a sonicación (UAE) durante 20 minutos a temperatura ambiente y se centrifugó durante 10 minutos a 3900 rpm. El sobrenadante se traspasó a un matraz vacío con el fin de evaporar el disolvente en el rotavapor a 38°C . El extracto final se obtuvo reconstituyendo en 2 mL de metanol/agua 80:20 (v/v). Por último, el extracto se filtró con un filtro de jeringa de $0.2\ \mu\text{m}$ (Millipore, Bedford, Massachusetts, EE.UU.) y se almacenó a -20°C hasta el momento de su análisis.

El proceso de extracción de las muestras de pimiento para su posterior análisis por HPLC-MS es muy parecido al anteriormente citado para preparar las muestras de frutas y vegetales. Se recolectó 1.5 kg de cada variedad de pimiento. Se procedió al triturado y liofilizado de la porción comestible, utilizando un liofilizador de Christ Alpha 1-2 LD (Shropshire, Reino Unido). Se tomaron 0.5 gramos de cada muestra, a los que se les añadieron 10 mL de metanol/agua 80:20 (v/v). El proceso de extracción durante 30 minutos fue



asistido por ultrasonidos y posteriormente la mezcla se centrifugó durante 15 minutos a 3500 rpm. El disolvente del sobrenadante se evaporó en el rotavapor y el residuo se reconstituyó en 0.5 mL de metanol/agua 80:20 (v/v). Finalmente, el extracto de cada muestra se filtró con un filtro de jeringa de 0.2 μm . Se prepararon tres réplicas de cada una de las variedades de pimiento. El metanol utilizado para disolver las muestras fue de Panreac (Barcelona, España) y el agua se obtuvo del sistema Milli-Q de Millipore (Bedford, Massachusetts, EE.UU.).

B.2) Extracción de compuestos volátiles y compuestos fenólicos de *Cistus*.

La extracción llevada a cabo para las muestras de *Cistus* es diferente, en función de los compuestos bioactivos objeto de estudio. En primer lugar, se explica el procedimiento de extracción realizado en el estudio de los compuestos volátiles de *Cistus* y, en segundo lugar, se explica el procedimiento realizado para la preparación los extractos polifenólicos.

- **Compuestos volátiles.**

Para el estudio de los volátiles, las muestras se procesaron en las 24 horas después de la recolección y se trituraron con un molinillo para obtener partículas de tamaño uniforme. La extracción se realizó mediante microextracción en fase sólida del espacio en cabeza (HS-SPME), provisto por el equipo utilizado para la separación y detección de los compuestos. Se trata de un cromatógrafo de gases Agilent 7890A acoplado a un sistema de detección Agilent 5975C y equipado con un sistema de inyección automatizada (CTC-Combipal). La fibra utilizada fue una fibra de doble recubrimiento, divinilbenceno (DVB)/carboxen en polidimetilsiloxano (PDMS) de Supelco (Bellefonte, Pensilvania, EE.UU.).



Se pesó una cantidad de 0.10 g de cada planta, que inmediatamente se colocó en un vial de SMPE de 15 mL. En cada extracción, la muestra se mantuvo durante 10 min para alcanzar el equilibrio térmico y seguidamente, la fibra se expuso al espacio de cabeza de la muestra durante 30 min. La muestra se mantuvo en agitación a 40 °C y el tiempo de desorción fue de 20 min en el inyector del cromatógrafo de gases a 230 °C.

- **Compuestos fenólicos.**

Para el estudio de los compuestos fenólicos de *Cistus*, el sistema de extracción usado fue sólido-líquido (SLE), utilizando distintos disolventes. Tras trituración, algunas muestras se sometieron a maceración acuosa, y otras a maceración hidroalcohólica (etanol: agua, 1:1) durante 4 horas con agitación suave a temperatura no superior a 65°C. Los extractos líquidos se filtraron para eliminar sólidos en suspensión y posteriormente, se llevaron al rotavapor para concentrarlas (o eliminar el etanol, en el caso de extracción hidroalcohólica) hasta obtener 8° Brix aproximadamente antes del secado. Los extractos se sometieron a dos procedimientos diferentes de secado: liofilización o "freeze drying" (FD; Telstar Cryodos 80, España) y spray drying (SDY; Buche Mini Spray Dryer B-290, España). Los extractos resultantes en polvo se disolvieron en el tampón apropiado y se filtraron con filtros de 0.2 µm. De este modo, se obtuvo en primer lugar un total de 12 extractos, teniendo en cuenta las combinaciones entre las cuatro especies de *Cistus*, los dos tipos de maceración y los dos tipos de secado (Tabla 2). Se seleccionó uno de los extractos (S2) para ser purificado mediante el proceso descrito previamente por Beltrán-Debón *et al.* (2010) y Herranz-López *et al.* (2012)^{128,129}. El fraccionado se llevó a cabo mediante

¹²⁸ Beltrán-Debón, R. *et al.*, 2010. The aqueous extract of *Hibiscus sabdariffa* calices modulates the production of monocyte chemoattractant protein-1 in humans. *Phytomedicine*, 17(3-4), pp.186–91.

¹²⁹ Herranz-López, M. *et al.*, 2012. Synergism of plant-derived polyphenols in adipogenesis: perspectives and implications. *Phytomedicine*, 19(3-4), pp.253–61.



cromatografía de afinidad con una resina Amberlite FPX66 de 2x25 cm (The Dow Chemical Company, Midland, Míchigan, EE.UU.), que muestra alta afinidad por los compuestos fenólicos. Para ello, el extracto S2 se disolvió en agua destilada y se centrifugó a 3000 rpm. Se recogió el sobrenadante para cargarlo en la columna Amberlite a un flujo de 4 mL/min. Después, se procedió al lavado con 30 volúmenes de agua destilada y a la elución utilizando 3 volúmenes de etanol a 2 mL/min. Tras este proceso, se obtuvieron dos extractos nuevos: la fracción retenida por la columna y eluída con etanol (correspondiente al extracto codificado como S2a) y la fracción lavada con agua destilada (extracto S2b), que se colectaron por separado y se liofilizaron para los análisis posteriores.

A la colección de extractos de *Cistus*, se incluyó un extracto más. Siguiendo las mismas condiciones del proceso de extracción realizado para obtener el extracto S2 (maceración acuosa y spray drying), se obtuvo el extracto denominado S4. La única diferencia entre los dos extractos es el tiempo de recolección (noviembre, en el caso de S2 y agosto-septiembre, en el caso de S4). Estos tres extractos incluidos en el estudio en último lugar se presentan a la derecha en la Tabla 2.



Tabla 2. Extractos polifenólicos de diferentes especies de *Cistus* obtenidos mediante distintos métodos de extracción y secado.

Especies de <i>Cistus</i>	Extracción/Secado	Extracto
<i>C. ladanifer</i>	H ₂ O/ FD	L1
	H ₂ O/ SDY	L2
	H ₂ O-EtOH/ FD	L3
<i>C. albidus</i>	H ₂ O/ FD	A1
	H ₂ O/ SDY	A2
	H ₂ O-EtOH/ FD	A3
<i>C. clusii</i>	H ₂ O/ FD	C1
	H ₂ O/ SDY	C2
	H ₂ O-EtOH/ FD	C3
<i>C. salviifolius</i>	H ₂ O/ FD	S1
	H ₂ O/ SDY ¹	S2
	H ₂ O-EtOH/ FD	S3

Extracción/Secado	Extracto
H ₂ O/ SDY ¹ fracción retenida	S2a
H ₂ O/ SDY ¹ fracción lavada	S2b
H ₂ O/ SDY ²	S4

H₂O (maceración acuosa).

H₂O-EtOH (maceración hidroalcohólica).

FD (lío-filización).

SDY (spray drying).

¹ Material vegetal recolectado en Noviembre.

² Material vegetal recolectado en Agosto-Septiembre.

S2a y S2b, procedentes del fraccionamiento de S2.

El estudio de los compuestos fenólicos de *Cistus*, además de incluir caracterización, incluyó el estudio de valoración de la capacidad antimicrobiana de los extractos. Con ambos datos, se realizó posteriormente el estudio de correlación entre datos de bioactividad y datos de MS para la identificación de compuestos candidatos a ser responsables de la bioactividad. Para este estudio, se incrementó el número de muestras hasta 26. Para ello, a los 15 extractos detallados anteriormente, se sumaron 11 extractos más mediante la inclusión de dos especies de *Cistus* y dos procedimientos de secado adicionales, distintos de los previamente descritos. Para este estudio de correlación, se mantiene el ciego, de manera que los códigos de las muestras van del 1 al 26.



Tal y como se ha especificado en el apartado A.2. de la metodología, referente a la selección de muestras de *Cistus*, las dos especies incluídas en último lugar fueron *C. psilosepalus* y *C. populifolius*. En cuanto a los métodos de secado de los extractos nuevamente incluidos en el estudio, éstos fueron: secado mediante horno de vacío y secado mediante tecnología D.I.C. (descompresión instantánea controlada) seguida de spray-drying. En la Figura 23 se presenta un esquema con todas las especies, los métodos de extracción, de secado y purificación, cuyas combinaciones dieron lugar a los 26 extractos

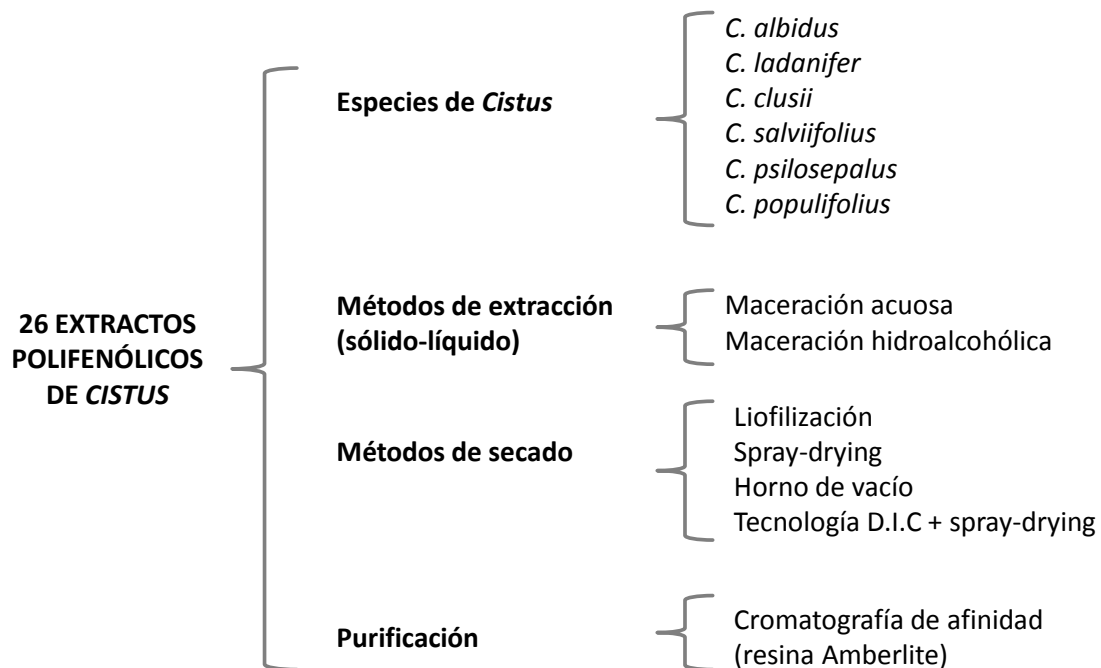


Figura 23. Colección de 26 extractos polifenólicos de *Cistus* para el estudio de correlación.

B.3) Extracción de compuestos fenólicos de aceite de oliva.

El sistema de extracción empleado para la preparación de extractos polifenólicos de aceite de oliva fue la extracción en fase sólida (SPE) mediante la utilización de cartuchos DIOL. Para llevarlo a cabo, primeramente, se pesaron 60 g de cada muestra de aceite, se diluyeron con 60 mL de hexano y la mezcla

resultante se homogeneizó en un agitador. Antes de hacer pasar la disolución preparada a través del cartucho, éste se acondicionó con 10 mL de metanol, seguidos de 10 mL de hexano. Al terminar de pasar la disolución de aceite y hexano, se lavó el cartucho con 15 mL de hexano. Para la elución del hexano, se utilizaron 40 mL de metanol. Después de la extracción, el extracto se llevó a sequedad con el rotavapor a una temperatura inferior a 40°C. Por último, se reconstituyó en 2 mL de metanol. Previo el análisis HPLC-MS, se preparó una dilución 1:10 de cada uno de los extractos. Para los ensayos de actividad anticancerígena, se realizó el mismo proceso de extracción con cada una de las muestras de aceite de oliva virgen extra, con la diferencia de ser reconstituidos en 1 mL de etanol tras la evaporación.

C) ANÁLISIS DE COMPOSICIÓN QUÍMICA.

C.1) Análisis y caracterización de compuestos fenólicos de pimiento.

En el estudio 1, se analizaron los extractos obtenidos con metanol/agua derivados de las tres variedades de pimiento incluidas en el estudio. Para la separación cromatográfica se utilizó un sistema Agilent 1200 RRLC (Agilent Technologies, Waldbronn, Alemania), equipado con un desgasificador de vacío, un inyector automático, una bomba binaria y un detector UV-vis. La separación se llevó a cabo con una columna analítica Zorbax Eclipse Plus C₁₈ (4.6 x 150 mm, 1.8 µm de tamaño de partícula), operando a 30°C y a un flujo de 0.8 mL/min. Las fases móviles que se utilizaron fueron agua con ácido acético (0.5%) (fase A) y acetonitrilo (fase B). El ácido acético y el acetonitrilo, ambos de grado analítico, fueron de Fisher Scientific (Leicestershire, Reino Unido) y Lab-Scan (Gliwice, Sowinskiego, Polonia), respectivamente. El agua utilizada para la preparación de la fase móvil A se obtuvo de un sistema de agua milli-Q de Millipore (Bedford, Massachusetts, EE.UU.). Una vez preparadas las fases



móviles, se filtraron con un sistema de filtración modelo 58.061 de Supelco (Bellefonte, Pensilvania, EE.UU.).

El gradiente utilizado siguió los pasos que se describen a continuación: 0 min, 0% B; 20 min, 20% B; 30 min, 30% B; 40 min, 50% B; 50 min, 75% B; 60 min, 100% B; 62 min, 0% B y finalmente, un ciclo de acondicionamiento de 8 minutos con las condiciones iniciales. Se inyectó un volumen de 10 μ L de extracto de pimiento, realizándose tres repeticiones de cada extracto.

El sistema RRLC se acopló a un microTOFTM (Bruker Daltonik, Bremen, Alemania) a través de una interfase ESI (modelo G1607A de Agilent Technologies, Palo Alto, CA, EEUU)(ver Figura 24). El sistema de ionización operó en modo negativo. Siendo el flujo utilizado en el método RRLC demasiado alto para lograr una ionización por electrospray estable, se precisó de la utilización de un divisor de flujo 1:3, de manera que el flujo suministrado al espectrómetro de masas fue de 0.2 mL/min. De acuerdo con esto, los valores óptimos de los parámetros de la fuente ESI-TOF fueron: voltaje a la entrada del capilar, ± 4.0 kV; temperatura del gas de secado, 190 °C; flujo del gas de secado, 8 L/min; presión del gas nebulizador, 2 bar; y end-plate offset, ± 0.5 kV. Los valores óptimos de los parámetros de transferencia fueron los siguientes: salida del capilar, ± 150 V; skimmer 1; ± 40 ; hexapolo 1; ± 23 V; RF hexapolo 50 Vpp; y skimmer 2, ± 22.5 V. Para la detección de los compuestos de interés, se consideró un rango de masas de 50 a 1000 m/z.





Figura 24. Equipo de cromatografía líquida de resolución rápida (RRLC) acoplado a espectrómetro de masas con analizador de tiempo de vuelo (TOF-MS) mediante ionización por electrospray (ESI).

Se realizó una calibración externa mediante el uso de una bomba de jeringa (Cole Palmer, Vernon Hills, Illinois, EE.UU.) directamente conectada a la interfase. El calibrante utilizado fue una disolución de acetato sódico, que se inyectó al comienzo de cada análisis. Este calibrante se preparó mezclando una solución de hidróxido sódico 1 M (Sigma-Aldrich, San Luis, Misuri, EE.UU.) y otra de agua: isopropanol 1:1 (v/v) (Lab-Scan, Gliwice, Sowinskiego, Polonia) con 0.2% de ácido acético (Fisher Scientific, Leicestershire, Reino Unido) a una relación 5:995 (v/v). Éste fue inyectado al comienzo de cada análisis para la identificación de los compuestos fenólicos. De este modo, se obtuvo una curva de calibración basada en numerosos *clusters* de masas, con una diferencia entre éstos de 68 Da (NaCHO_2).

Los datos de masas exactas de los iones analizados se procesaron con el software DataAnalysis 3.4 (Bruker Daltonik), el cual es capaz de ofrecer una lista de posibles fórmulas moleculares empleando el editor GenerateMolecularFormula™. Este editor utiliza un algoritmo CHNO, que posee una serie de funcionalidades como el rango elemental mínimo/máximo, la configuración electrónica, equivalentes de dobles enlaces y anillos, así como una comparación del patrón isotópico teórico y experimental (*mSigma*) para conseguir una mayor seguridad en la fórmula molecular propuesta. El umbral de exactitud más ampliamente aceptado por la comunidad científica para la confirmación de la composición elemental se establece en 5 ppm. La identificación de los compuestos se llevó a cabo mediante la comparación del tiempo de retención y los datos MS (masa exacta y distribución isotópica) de los compuestos detectados con los compuestos de pimienta u otros vegetales de la misma familia publicados en bibliografía y la búsqueda en bases de datos on-line como SciFinder Scholar (<<https://scifinder.cas.org/>>) y Phenol-Explorer (<<http://www.phenol-explorer.eu/>>).

C.2) Análisis y caracterización compuestos volátiles y compuestos fenólicos de *Cistus*.

Las muestras de *Cistus* se analizaron mediante tres procedimientos diferentes. Por un lado, se utilizó cromatografía de gases acoplada a espectrometría de masas (GC-MS) para el análisis de volátiles y por otro lado, se utilizó cromatografía líquida acoplada a espectrometría de masas (LC-MS) para el análisis de compuestos fenólicos. Dentro de las técnicas de LC-MS, se utilizaron distintas plataformas analíticas: cromatografía líquida de alta resolución (HPLC) acoplada a MS con analizador de tiempo de vuelo (TOF), y cromatografía líquida de ultra alta resolución (UPLC) acoplada a MS con analizador de cuadrupolo-tiempo de vuelo (QTOF).



- **Compuestos volátiles.**

El análisis de compuestos volátiles se realizó por desorción térmica en el inyector del cromatógrafo de gases de los compuestos, previamente extraídos por la fibra en el espacio de cabeza del vial de microextracción en fase sólida. Para la separación cromatográfica, se utilizó una columna capilar HP-5 de Agilent (30m x 250 μm x 1 μm). El programa de temperatura empleado fue: 5 min isotérmico a 40 $^{\circ}\text{C}$ y rampa hasta 230 $^{\circ}\text{C}$ a 4 $^{\circ}\text{C}/\text{min}$. Se utilizó helio a 2,1 mL/min como gas portador. La temperatura del inyector, la interfase y el detector fue de 250 $^{\circ}\text{C}$. La muestra se inyectó en modo *splitless* (sin división de muestra). Los espectros de masas se midieron en modo de impacto electrónico con un voltaje de ionización de 70 eV, siendo el rango de *scan* entre 20-280 m/z

Se preparó una mezcla patrón de alcanos C6–C25 a una concentración de 100 $\mu\text{g}/\text{mL}$ en dimetiléter (Sigma–Aldrich, Europa) que se analizó bajo las mismas condiciones que las muestras. Teniendo en cuenta los tiempos de retención obtenidos para los patrones, se calcularon los índices de retención de Kovats de los compuestos de las muestras.

Para la caracterización, se utilizó el software ChemStation de Agilent. Los compuestos fueron identificados mediante comparación con espectros de masas procedentes de las bases de datos NIST/EPA/NIH y confirmados en muchos casos por comparación de sus índices de retención con bases de datos (<http://www.pherobase.com>), así como la librería ADAMS que se utiliza principalmente para terpenos. Las determinaciones cuantitativas se expresaron como porcentaje de área relativa a la suma de las áreas de la totalidad de compuestos identificados.



- **Compuestos fenólicos.**

El análisis de compuestos fenólicos de *Cistus* se llevó a cabo utilizando un método similar al descrito previamente en otros estudios de esta misma planta¹³⁰. La separación analítica y la detección de los compuestos se llevó a cabo empleando el equipamiento descrito en el apartado C.1 sobre el análisis de las muestras de pimiento, es decir, un sistema RRLC acoplado a un TOF con interfase ESI. La columna analítica utilizada también fue una C₁₈. Las fases móviles consistieron en 0.5% de ácido acético (A) y acetonitrilo (B), siendo de Fisher Scientific (Leicestershire, Reino Unido) y Lab-Scan (Gliwice, Sowinskiego, Polonia), respectivamente. El agua doblemente ionizada de baja conductividad se obtuvo de un sistema de agua milli-Q de Millipore (Bedford, Massachusetts, EE.UU.). Las fases móviles se filtraron con un sistema de filtración modelo 58.061 de Supelco (Bellefonte, Pensilvania, EE.UU.).

El gradiente que se aplicó fue: 0 min, 0% B; 20 min, 20% B; 30 min, 30% B; 40 min, 50% B; 50 min, 75% B; 60 min, 100% B. El volumen de inyección fue de 15 µL, trabajando a un flujo de 0.8 mL/min. Se utilizó un divisor de flujo para el acoplamiento RRLC-MS, con el que se obtuvo un flujo de entrada al espectrómetro de 0.25 mL/min.

Para la calibración externa del TOF se utilizaron *clusters* de acetato sódico como calibrante. Los parámetros de la fuente de ionización operando en modo negativo fueron los siguientes: presión del gas nebulizador, 2 bar; flujo del gas nebulizador, 9 L/min y temperatura del gas de secado, 190 °C. El voltaje en la cámara de nebulización fue de 500 V y el voltaje de entrada al capilar fue de

¹³⁰ Fernández-Arroyo, S. *et al.*, 2010. High-performance liquid chromatography with diode array detection coupled to electrospray time-of-flight and ion-trap tandem mass spectrometry to identify phenolic compounds from a *Cistus ladanifer* aqueous extract. *Phytochemical Analysis*, 21(4), pp.307–313.



4500 V. El voltaje utilizado en el tubo de vuelo fue de +9000 V, un voltaje de +1960 V en el detector, considerando un rango de masas de 50 a 1100 m/z.

Para la caracterización de todos extractos polifenólicos de *Cistus* especificados en el apartado B.2, se utilizó el software DataAnalysis 4.0 (Bruker Daltonik), una versión más actualizada del software utilizado en el apartado anterior, pero con la misma funcionalidad, ya que ofrece una lista de posibles fórmulas moleculares usando el editor SmartFormula™.

Posteriormente, en base a los resultados de bioactividad, se seleccionaron tres de los extractos para estudiar su composición fenólica más pormenorizadamente: S2, S2a y S4 (procedencia y obtención de los extractos se detalla en el apartado B.2 de la metodología). En este caso, el análisis se llevó a cabo mediante UPLC–ESI–QTOF–MS (ver Figura 25).



Figura 25. Equipo de UPLC Acquity Waters y microTOF Q-II de Bruker Daltonik.

Para estos tres extractos de *Cistus*, la separación cromatográfica se llevó a cabo con un equipo UPLC de Waters ACQUITY, utilizando una columna analítica Zorbax Eclipse Plus C₁₈ (1.8 μm, 150×4.6 mm). Las fases móviles utilizadas consistieron en agua acidificada con ácido acético al 0.5% (fase A) y acetonitrilo (fase B). El ácido acético y el acetonitrilo de grado analítico fueron obtenidos de

Fluka (Sigma-Aldrich, Steinheim, Alemania) y Lab-Scan (Gliwice, Sowinskiego, Polonia), respectivamente. Para la fase móvil A se utilizó agua obtenida de un sistema Milli-Q (Millipore, Bedford, Massachusetts, EE.UU.). Las fases móviles se filtraron con un sistema de filtración modelo 58.061 (Supelco, Bellefonte, Pensilvania, EE.UU.). El gradiente utilizado consistió en los siguientes pasos: 0 min, 0% B; 20 min, 20% B; 30 min, 30% B; 40 min, 50% B; 50 min, 75% B; 60 min, 100% B; 62 min, 0% B. Las condiciones iniciales se mantuvieron durante 8 min.

El flujo con el que se trabajó en el UPLC fue de 0.8 mL/min, inyectando un volumen de 7 μ L. Este equipo se acopló a un espectrómetro de masas con analizador de cuadrupolo-tiempo de vuelo microTOF-Q™ (Bruker Daltonik, Bremen, Alemania), equipado con un sistema de ionización por electrospray operando en modo negativo. El flujo introducido al espectrómetro fue de 0.2 mL/min, tras el uso de un divisor de flujo 1:4.

La adquisición de datos tuvo lugar considerando un rango de masas de m/z 50–1500, algo mayor que en el estudio previo referente al resto de extractos polifenólicos de *Cistus*, con el fin de poder detectar taninos hidrolizables de alto peso molecular como galotaninos y elagitaninos.

Los parámetros de la fuente de ionización y la zona de transferencia se detallan a continuación: voltaje del capilar, +3500 V; temperatura del gas de secado, 200 °C; flujo del gas nebulizador, 8.0 L/min; presión del gas nebulizador, 2.0 bar. Se realizaron análisis de MS/MS automáticos, ajustando los valores de energía de colisión como se especifica a continuación: m/z 100, 20 eV; m/z 500, 25 eV; m/z 1000, 30 eV, y utilizando nitrógeno como gas de colisión. Como calibrante, se utilizó una solución de acetato sódico, la cual se inyectó al principio de cada análisis mediante una jeringa Cole Palmer (Vernon Hills, Illinois, EE.UU.) conectada a la interfase del QTOF.



En cuanto a la identificación de los compuestos proporcionados en la lista, se utilizó el software DataAnalysis 4.0 (Bruker Daltonik) y su editor SmartFormula™. Los compuestos fueron identificados en base a los espectros de MS y MS/MS registrados, la información disponible en bibliografía y haciendo uso de las bases de datos.

C.3) Análisis de compuestos fenólicos de aceite de oliva.

El análisis de los extractos polifenólicos de aceite de oliva virgen extra se realizó mediante la utilización de un método HPLC–ESI–TOF–MS. El equipamiento utilizado fue el mismo que el utilizado para el análisis de las muestras de pimiento: un cromatógrafo Agilent 1200-RRLC acoplado a un espectrómetro de masas de tiempo de vuelo microTOF™, mediante un sistema de ionización por electrospray. La columna analítica utilizada fue una Zorbax Eclipse Plus C₁₈ (4.6 x 150 mm, 1.8 µm de tamaño de partícula). Como fases móviles A y B se emplearon agua ácida con un 0.25% de acético y metanol, respectivamente. El metanol se obtuvo de Lab-Scan (Dublín, Irlanda) y el ácido acético, de Sigma-Aldrich (Steinheim, Alemania). El agua utilizada para la fase móvil A se obtuvo de un sistema Milli-Q (Millipore, Bedford, Massachusetts, EE.UU.). Antes de su uso, ambas fases se filtraron con un sistema de filtración modelo 58.061 (Supelco, Bellefonte, Pensilvania, EE.UU.).

El gradiente multietapa empleado fue el siguiente: 0 min, 5% B; 7 min, 35% B; 12 min, 45% B; 17 min, 50% B; 22 min, 60% B; 25 min, 95% B, 27 min, 5% B, y finalmente un ciclo de acondicionamiento de 5 min con las condiciones iniciales para el siguiente análisis. El flujo utilizado fue de 0.8 mL/min y la temperatura de la columna se mantuvo a 25°C. El volumen de inyección fue de 10 µL.



La interfase ESI utilizada entre ambos equipos operó en modo negativo, utilizando previamente un divisor de flujo 1:3, de manera que a la entrada del espectrómetro de masas el flujo se redujo de 0.8 a 0.24 mL/min.

Se llevó a cabo una calibración externa del microTOFTM utilizando clusters de formiato sódico como calibrante, que fue inyectado al comienzo de cada análisis. Los parámetros de la fuente fueron establecidos en función del flujo de entrada. Así, para obtener presiones de alto vacío constantes, se fijaron los siguientes valores: presión del gas nebulizador, 2 bar; flujo del gas nebulizador, 9 L/min y temperatura del gas de secado, 190°C. En cuanto al voltaje, en la cámara de nebulización y a la entrada el capilar, los valores fueron de -0.5 kV y 4 kV respectivamente. Los parámetros de la zona de transmisión o transferencia óptica fueron los siguientes: voltaje a la salida del capilar, -120 V; skimmer 1, -40 V; skimmer 2, -22.5V; hexapolo 1, -23V y radiofrecuencia del hexapolo 50 Vpp.

Para la caracterización de los compuestos, los datos se procesaron utilizando el software DataAnalysis 3.4 (Bruker Daltonik GmbH, Bremen, Alemania) y el editor GenerateMolecularFormulaTM. La identificación de los compuestos fenólicos fue llevada a cabo por comparación de los tiempos de retención y los espectros de masas de las muestras y los siguientes estándares comerciales: hidroxitirosol, tirosol, vanilina, luteolina, apigenina, ácidos *p*-coumárico, ferúlico y vanílico que fueron suministrados por Sigma-Aldrich, (+)-pinoresinol de Arbo Nova (Turku, Finland) y oleuropeína, que fue adquirida de Extrasynthèse (Lyon, Francia). Aquellos compuestos para los que no hay estándares comerciales disponibles fueron identificados utilizando la información disponible en bibliografía, así como mediante la interpretación de sus espectros de masas generados por el TOF-MS.



D) ANÁLISIS DE BIOACTIVIDAD.

D.1) Evaluación de capacidad antioxidante.

Las determinaciones de capacidad antioxidante en extractos de frutas y vegetales se llevaron a cabo con el Lector Multi-Modal de microplacas basado en monocromadores Synergy™ Mx de Biotek (Winooski, Vermont, EE.UU.) (Figura 26). Se evaluó mediante tres métodos diferentes de capacidad antioxidante: TEAC (trolox equivalent antioxidant capacity), FRAP (ferric ion reducing antioxidant power) y ORAC (Oxygen Radical Absorbance Capacity).



Figura 26. Lector de microplacas Synergy™ Mx BioTek y microplaca de 96 pocillos para valorar la capacidad antioxidante.

Los reactivos utilizados para los análisis de capacidad antioxidante, AAPH (2,2'-Azobis-(2-aminopropano)-dihidrocloruro), TPTZ (tripirydyltriazine), ABTS (2,2'-azinobis (3-etilbenzotiazolín)-6-sulfónico), Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid), fluoresceína (3',6'-dihidroxiestero-(2-benzofuran-3,9'-xanten)-1-ona), persulfato potásico ($K_2S_2O_8$) y sulfato férrico ($FeSO_4$) fueron de Sigma-Aldrich (San Luis, Misuri, EE.UU.). El fosfato monosódico (NaH_2PO_4), fosfato disódico (Na_2HPO_4), el cloruro férrico (Cl_3Fe), el ácido clorhídrico (HCl) y el ácido acético (CH_3COOH) se obtuvieron de Panreac (Barcelona, España). El agua milli-Q utilizada en la preparación de los reactivos

fue proporcionada por un sistema de Millipore (Bedford, Massachusetts, EE.UU.).

Se utilizó el programa Origin (versión Origin Pro 8 SRO, Northampton, Massachusetts, EE.UU.) para el cálculo de la media y la desviación estándar de los ensayos llevados cabo por triplicado.

A continuación, se detalla por separado la metodología de cada uno de los ensayos de capacidad antioxidante realizados para analizar las muestras de frutas y vegetales.

- **Método TEAC (trolox equivalent antioxidant capacity).**

El radical catión se preparó mezclando una disolución 7mM de ABTS en agua milli-Q con otra disolución de persulfato potásico 2.45 mM. Esta mezcla se dejó incubando durante 12-24 horas en oscuridad a temperatura ambiente. Posteriormente, dicha mezcla se diluyó en agua, para llevar a cabo el ensayo hidrofílico, hasta obtener una absorbancia de 0.70 ± 0.02 a una λ de 734 nm en el lector de placas.

Para construir la recta patrón, se prepararon distintas concentraciones de Trolox a partir de una disolución de Trolox 1 mM (correspondiendo a 0.5–30 μ M en los pocillos de la placa). Se añadieron 300 μ L de la disolución de ABTS^{•+} ajustada a todos los pocillos de la placa y se hizo una primera lectura a 734 nm, siendo éstos los datos de absorbancia inicial. A continuación, se añadieron 30 μ L, por triplicado, de las distintas disoluciones de Trolox y de las muestras. El blanco se hizo con agua milli-Q. Se introdujo la placa de nuevo en el lector, donde se agitó y se esperaron 3 min para que la reacción tuviera lugar. Se midió la absorbancia a 734 nm para obtener los datos de absorbancia final. El valor TEAC se expresó en mmol equivalentes de Trolox por 100 g de muestra (peso seco).



- **Método FRAP (ferric ion reducing antioxidant power).**

En primer lugar, se preparó una disolución con acetato de sodio 300 mM (1,6% de ácido acético, pH 3.6), tripyridyltriazine (TPTZ) 10 mM en 40 mM de ácido clorhídrico (HCl) y cloruro férrico (Cl_3Fe) 20 mM, mezclándolos a una relación 10:1:1.

El lector de placas se acondicionó a 37°C durante 5 min. Se añadieron 250 μL de la mezcla anterior a cada pocillo de una placa de 96 pocillos y se realizó una primera medida de absorbancia a 593 nm como datos de absorbancia inicial. A continuación se añadió, por triplicado, 40 μL de extracto vegetal convenientemente diluido, así como del patrón de sulfato férrico (FeSO_4) para realizar la recta de calibrado. Las concentraciones de patrón en el pocillo fueron de 12.5 a 200 μM . Como blanco se utilizó 40 μL de agua milli-Q. Se introdujo nuevamente la placa en el lector, donde se agitó y se esperaron 4 min para que la reacción tuviera lugar. Finalmente, se midió la absorbancia a 593 nm para obtener los valores de absorbancia final. El valor FRAP se expresó en mmoles equivalentes de FeSO_4 por 100 g de muestra (peso seco).

- **Método ORAC (Oxigen Radical Absorbance Capacity).**

Se preparó en tampón ORAC (fosfato 75 mM a pH 7.4) mediante una disolución 80:20 de fosfato disódico (Na_2HPO_4) y fosfato monosódico (NaH_2PO_4). A partir de este tampón, se preparó la fluoresceína 4.53 mM. En una placa de 96 pocillos, se añadió a cada uno 150 μL de fluoresceína, siendo la concentración en el pocillo de 40 nM. La fluoresceína se incubó previamente durante un mínimo de 30 min a 37 °C en estufa. Se añadieron 30 μL de las distintas disoluciones de extracto y de patrón (0.5-15 μM Trolox/pocillo), por triplicado, y se dejó atemperar dentro del lector 15 min a 37 °C. Se preparó un control negativo (C0) en el que no existe decaimiento de la fluorescencia, añadiendo a



los 150 μ L de ésta, 60 μ L de tampón, así como un control positivo (C60) en el que la fluorescencia decae prácticamente al instante, añadiendo 30 μ L de tampón y 30 μ L de AAPH. Una vez transcurridos los 15 min, se añadió a cada pocillo (excepto al C0) 30 μ L de AAPH 133 mM, disolución preparada en tampón justo antes de su uso.

Los datos del decaimiento de la intensidad de fluorescencia se registraron cada 60 s durante 3 h a 37 °C con agitación orbital antes de cada medición, a una longitud de onda de excitación y emisión de 485 y 520 nm, respectivamente e inmediatamente después de agregar el AAPH. Los valores finales de ORAC se calcularon mediante una ecuación de regresión entre la concentración de Trolox y el área neta bajo la curva de la fluorescencia (area under curve, AUC). El valor ORAC se expresó en mmol equivalentes Trolox por 100 g de muestra (peso seco).

D.2) Evaluación de capacidad antimicrobiana.

Para evaluar la capacidad antimicrobiana de los extractos polifenólicos de *Cistus*, se realizó una serie de ensayos de inhibición del crecimiento bacteriano, en colaboración con el IBMC de la Universidad Miguel Hernández.

Para ello, se utilizaron cepas de *Escherichia coli* (CECT 515) y *Staphylococcus aureus* (CECT 59) como modelos de bacterias Gram negativa y Gram positiva, respectivamente. Ambas cepas fueron obtenidas de la Colección Española de Cultivos Tipo, CECT, Universitat de Valencia, España). La determinación de los valores de MIC₅₀ se desarrolló en placas de 96 pocillos mediante un ensayo de susceptibilidad de microdilución en caldo, desarrollado mediante modificaciones de un método previamente descrito¹³¹.

¹³¹ İşcan, G. *et al.*, 2002. Antimicrobial screening of *Mentha piperita* essential oils. *Journal of Agricultural and Food Chemistry*, 50(14), pp.3943–3946.



En primer lugar, se dejó crecer un preinóculo de 5 mL durante la noche a 37°C en caldo LB (Conda Pronadisa, Europa). De ahí, se preparó un inóculo mediante dilución del 10% del preinóculo en el mismo medio, y éste se incubó durante 4 horas a 37°C con el fin de asegurar un crecimiento exponencial. Después de esta incubación, se hicieron diluciones seriadas de una alícuota del cultivo hasta obtener una concentración bacteriana de 10^3 ufc/mL. A cada pocillo de la microplaca se añadieron 20 μ L del cultivo diluido más 100 μ L de medio fresco LB y 80 μ L del extracto disuelto en PBS. Los extractos se diluyeron en un tampón PBS para obtener una serie de concentraciones de entre 2,000 y 3.9 μ g/mL. El ensayo de susceptibilidad se realizó para cada concentración por octuplicado.

Como control positivo, se utilizó Neomicina (Sigma-Aldrich, Europa) en cada ensayo. Como controles negativos, se utilizaron las cantidades apropiadas de medio LB fresco, de disolvente en el que los extractos fueron diluidos y de las soluciones que contenían los extractos, pero no se utilizaron bacterias. Se incubaron las microplacas a 37°C durante 16 horas para *E. coli* y 24 horas para *S. aureus*.

La supervivencia microbiana después de los tratamientos se midió colorimétricamente utilizando el p-iodonitrotetrazolium (INT) (Sigma-Aldrich, Europa). Se añadieron 50 μ L de INT a cada pocillo para llegar a una concentración final de 0.25 mg/mL y se incubó 30 minutos a 37°C. Finalmente, se midió la señal de color rojo resultante a 570 nm mediante el uso de un espectrofotómetro (SPECTROstar Omega, BMG Labtech GmbH, Offenburg, Alemania).

Para procesar los resultados, se utilizó un software Graphad Prism (version 5.0). Los valores MIC₅₀ (concentración que inhibe el 50% de la actividad bacteriana) se obtuvieron aplicando una curva sigmoideal de pendiente variable ajustada a los datos experimentales y obteniendo la curva de dosis-respuesta. El intervalo



de confianza del 95% y el ajuste de la curva (R^2) fue también proporcionado por el software para todas las muestras. A partir de los valores de MIC_{50} de tres experimentos independientes, se obtuvo la media y la desviación estándar.

D.3) Evaluación de capacidad anticancerígena.

Se determinó la capacidad anticancerígena de los extractos polifenólicos de aceite de oliva virgen extra (extra virgin olive oil phenolic extract, EVOO-PE) frente a la línea celular de cáncer de mama JIMT-1. Para ello, se realizaron una serie de ensayos de viabilidad celular, en colaboración con en el Instituto Catalán de Oncología (ICO) de Girona. Estos bioensayos se basaron en la utilización de MTT (bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolium).

El cultivo de las células JIMT-1, disponible en la Colección alemana de Microorganismos y Cultivos Celulares (<http://www.dsmz.de/>), se realizó en medio DEMEN/F-12 (1:1) suplementado con suero fetal bovino al 10 % y 2 mM de L-glutamina. Las células se mantuvieron a 37°C en una atmósfera humidificada de 95% de aire y 5% de CO₂.

En primer lugar, se sembraron las células JIMT-1 a una alta densidad (3000 células por pocillo en una placa de 96 pocillos). Al día siguiente, se trataron con concentraciones crecientes de los extractos de EVOO disuelto en etanol, desde 0% hasta 0,1% (v/v). Para los controles sólo se añadió la cantidad adecuada de etanol. Tras 5 días de tratamiento, las células se incubaron con una solución de MTT (Sigma-Aldrich, San Luis, Misuri, EE.UU.) a una concentración de 5 mg/mL durante 3 h a 37 °C. De ahí, se aspiró el sobrenadante cuidadosamente, al cual se le añadieron 100 µL de DMSO en cada pocillo y las placas se agitaron para disolver el producto cristalino obtenido.

Se leyeron los valores de absorbancia a 570 nm utilizando un lector de placas multipocillo (modelo Anthos Labtec 2010 1.7). Los efectos de viabilidad de la



exposición de las células a cada extracto se analizaron como porcentaje de los valores de absorbancia de los controles. Es decir, que para cada tratamiento, la viabilidad celular se evaluó como porcentaje según la siguiente ecuación: $(A_{570}$ de la muestra tratada/ A_{570} de las muestras no tratadas) x 100. La sensibilidad de las células JIMT-1 a los extractos crudos de EVOO fue expresada en términos de concentración de extracto (v/v) requerida para disminuir en un 50% la viabilidad celular (valor IC_{50}). Asimismo, el porcentaje de absorbancia para los controles se consideró la fracción de supervivencia de las células, por lo que los valores de IC_{50} se definieron como la concentración de los extractos que produjo una reducción del 50% en la absorbancia control (por interpolación).

E) TRATAMIENTO DE LOS DATOS DE CROMATOGRAFÍA ACOPLADA A ESPECTROMETRÍA DE MASAS (LC-MS).

En la metodología seguida para la identificación de cada uno de los componentes de una muestra, se ha visto la necesidad de la utilización de una técnica separativa acoplada a un sistema de detección. En la Tabla 3, se presentan de manera esquemática las técnicas analíticas utilizadas en la presente tesis para el estudio de compuestos fenólicos. Concretamente, se ha utilizado como técnica separativa, la cromatografía líquida de alta resolución (HPLC), siendo acoplada mediante una interfase de ionización por electrospray (ESI) a un espectrómetro de masas de tiempo de vuelo (TOF) o cuadrupolo-tiempo de vuelo (Q-TOF). En adelante, se utilizara de manera simplificada el término LC-MS, en lugar de HPLC-ESI-(Q)TOF-MS.



Tabla 3. Esquema de las plataformas analíticas utilizadas para el análisis de compuestos fenólicos en la presente tesis: cromatografía líquida y espectrometría de masas.

Técnica separativa	Sistema de detección		
Cromatografía líquida de alta resolución (HPLC)	Espectrometría de masas (MS)		
	Sistema de ionización	Analizador de masas	Detector
	Electrospray (ESI)	Tiempo de vuelo (TOF)	Multiplicador de electrones
		Cuadrupolo-tiempo de vuelo (QTOF)	

Este apartado de metodología está dividido en dos subapartados. En el primero, se describe primeramente el fundamento de las técnicas de cromatografía y espectrometría de masas y las características de las señales registradas en el análisis de muestras simples y complejas mediante la utilización de dichas técnicas. Desde un punto de vista de teoría de la señal, se describen los procesos que afectan a los iones cuando vuelan dentro del espectrómetro de masas, así como su efecto sobre los registros que se recogen en el detector, con el fin de garantizar una mejor comprensión del tipo de tratamiento que es necesario realizar con los datos obtenidos. También se describe el tratamiento convencional de esos registros para dar paso en el segundo subapartado a la descripción detallada de la nueva estrategia de tratamiento de datos LC-MS desarrollada en esta tesis.



E.1) Fundamentos LC-MS desde un punto de vista de teoría de la señal.

- **Fundamentos de cromatografía líquida (LC).**

Como se ha descrito en la introducción, la LC es una técnica que permite la separación de los componentes de una muestra debido a la influencia de dos efectos contrapuestos:

- **Retención:** efecto producido sobre los componentes de la muestra por una fase estacionaria, que está constituida por un sólido apolar poroso, finamente granulado y que se deposita en el interior de una columna.
- **Desplazamiento:** efecto ejercido sobre los componentes de la muestra por una fase móvil, que está constituida por un disolvente (o mezcla de varios) en el que los componentes de la muestra deben ser solubles.

El fenómeno de migración de los componentes de una muestra a lo largo de la fase estacionaria, recibe el nombre de **elución**. La muestra, impulsada por la fase móvil, atraviesa la fase estacionaria, desplazando a los componentes de la muestra a distinta velocidad, dependiendo de la magnitud de sus interacciones relativas con ambas fases. Las dos fases se eligen de forma que los componentes de la muestra se distribuyan de modo distinto entre la fase móvil y la fase estacionaria. Aquellos componentes que son fuertemente retenidos por la fase estacionaria se mueven lentamente con el flujo de la fase móvil; por el contrario los componentes que se unen débilmente a la fase estacionaria, se mueven con rapidez. Como consecuencia de la distinta movilidad, los componentes de la muestra se separan en bandas o zonas discretas que pueden analizarse cualitativa y/o cuantitativamente.



La afinidad de los componentes de la muestra por la fase estacionaria y la fase móvil dependerá de su polaridad. En el caso de los compuestos fenólicos, por su carácter polar, se utiliza cromatografía de partición en fase inversa, en la que la fase estacionaria es apolar y la fase móvil es polar.

El resultado gráfico de la cromatografía se denomina **cromatograma**, en el que se observa la distribución final de los componentes de la muestra a lo largo del tiempo de análisis. El tiempo necesario para eluir una molécula de la columna se llama **tiempo de elución o de retención (t_R)**^{82,83}.

- **Forma de los picos cromatográficos: dispersión de bandas.**

Las bandas cromatográficas se van ensanchando al mismo tiempo que se separan; este proceso (de difusión molecular) es debido fundamentalmente a la difusión termodinámica de las moléculas de la banda, que tienden a distribuirse uniformemente en todo el volumen disponible. Dado que los procesos de difusión de las moléculas tienen su fundamento en movimientos de las mismas al azar, la concentración final de las moléculas dentro de una banda cromatográfica adoptará, en su caso ideal, un perfil de distribución normal, y el registro de las concentraciones de la banda en función del volumen eluido o del tiempo de retención, dará lugar a una forma gaussiana de anchura definida por la desviación estándar de la dispersión.

Para realizar una separación cromatográfica óptima, debe evitarse en la medida de lo posible, el ensanchamiento de las bandas debido a la dispersión; cuanto más anchas sean las bandas, menor número de ellas podrán ser resueltas en un intervalo de tiempo dado. Los cromatogramas deben contener picos resueltos en forma gaussiana lo más estrechos posible. Son muchos los parámetros que afectan a la separación cromatográfica y sólo se puede actuar sobre algunos para optimizar el método cromatográfico.



Los tiempos de retención y la anchura y separación de los picos cromatográficos dependerán principalmente de la polaridad de los compuestos a separar, de la naturaleza de la fase estacionaria (polaridad, selectividad, tamaño de partícula,..), del diámetro interno y longitud la columna y de la naturaleza de la fase móvil (de los disolventes que la componen, de la velocidad de flujo,...). Todos estos parámetros afectan a la resolución, rapidez y sensibilidad de la separación cromatográfica.

- **Acoplamiento LC-MS.**

El acoplamiento de un equipo cromatográfico con un espectrómetro de masas es la técnica combinada más ampliamente utilizada ya que reúne la enorme capacidad de separación que ofrece el cromatógrafo, con la sensibilidad y capacidad de aportar información estructural del espectrómetro de masas.

El principal problema que se plantea a la hora de acoplar estas dos técnicas es que la cromatografía, trabaja a presión superior a la atmosférica, mientras que el espectrómetro de masas trabaja con alto vacío. Éste y otros inconvenientes, se solucionan utilizando entre los dos equipos un dispositivo de conexión intermedio o interfase, como el utilizado en esta tesis, de ionización por electrospray (ESI).

Los requerimientos que deben cumplir los espectrómetros de masas para acoplarse a una técnica cromatográfica son fundamentalmente dos: ser capaces de trabajar con las pequeñas cantidades de componentes de la muestra, que eluyen de la columna cromatográfica (normalmente del rango de 1 μg o 1 ng) y que el tiempo de adquisición de datos sea rápido para que pueda recoger en tiempo real todos los compuestos que van eluyendo de la columna (velocidad típica de un análisis).



- **Fundamentos de los espectrómetros de masas (MS) de tiempo de vuelo (TOF).**

La espectrometría de masas proporciona una información muy útil para el estudio cualitativo y cuantitativo de mezclas complejas. Como resultado del análisis LC-MS, se obtiene un gran volumen de datos, en el que los distintos compuestos aparecen separados por tiempos de retención y por relación masa/carga (m/z), proporcionando, el instrumento de medida, la intensidad (o corriente iónica) para las distintos valores de tiempo y relación masa/carga. Los equipos de espectrometría de masas basados en tiempo de vuelo se caracterizan, además, por proporcionar una precisión muy alta en la medida de la masa de los iones.

Como se ha mencionado en la introducción (apartado 6), las partes básicas del espectrómetros de masas son: un sistema de introducción de muestra, un sistema de vacío, un medio por el que las muestras son ionizadas, un acelerador de iones mediante un campo eléctrico, un sistema de dispersión de iones de acuerdo a su relación m/z y un detector de los iones que convierte el haz de iones en una señal eléctrica que puede ser entonces procesada, almacenada en la memoria de un ordenador y mostrada o registrada de varias maneras.

El fundamento del analizador de **tiempo de vuelo** (*time-of-flight*, TOF) se basa en la medición del tiempo requerido por un ión para viajar desde la fuente de iones hasta el detector localizado a una cierta distancia de la fuente.

Los iones son extraídos de la fuente de ionización en pulsos e inmediatamente son acelerados mediante la acción de un campo eléctrico para lograr la uniformidad de sus energías cinéticas antes de que entren al analizador. Esto proporciona a todos los iones del paquete una misma energía cinética definida por la Ecuación 1:



$$\frac{1}{2}mv^2 = qV \quad (\text{Ecuación 1})$$

donde, m es la masa del ión, q es la carga del ión ($q=ze$, número de cargas de un ión por la carga de un electrón), v es la velocidad final alcanzada en la aceleración, $Ep=qV$ es la energía potencial de una partícula de carga q , y V es el potencial eléctrico aplicado. Posteriormente, los iones en movimiento rectilíneo uniforme atraviesan el tubo de vuelo, sometido a alto vacío, hacia el detector con una velocidad dada por la Ecuación 2:

$$v = \sqrt{2eV \frac{z}{m}} \quad (\text{Ecuación 2})$$

Es decir, la velocidad adquirida por cada ión disminuye con la raíz de su relación m/z . Los iones con valores m/z altos viajarán más lentos que los de m/z baja. El tiempo necesario para atravesar el tubo desde la fuente de iones hasta el detector (separados por una distancia L) o tiempo de vuelo (TOF) depende de la razón m/z de cada ión y viene dado por la Ecuación 3:

$$TOF = \frac{L}{v} = L \sqrt{\frac{m}{z} 2eV} = k \cdot \sqrt{\frac{m}{z}} \quad (\text{Ecuación 3})$$

donde k es una constante que recoge los parámetros de diseño del analizador (L fijo, V constante). De este modo, se puede determinar la masa (o razón m/z) para cada ión midiendo su tiempo de vuelo (TOF) al detector. Los valores de tiempo de vuelo se encuentran en torno a unos pocos microsegundos: por ejemplo, si se aplica un potencial de 3000 V, un ión de $m/z=800$ Da tarda 36 μ s en atravesar un tubo de vuelo de 1 m de longitud. Esto exige unos sistemas de detección extremadamente rápidos⁸⁹.

Los iones que parten de la fuente de ionización no son generados en un punto geométrico, sino en un volumen de espacio, y no tienen los mismos tiempos de partida ni exactamente la misma energía cinética (o velocidad inicial). Esto hace



que las partículas individuales correspondientes a un ión no lleguen al detector en un mismo instante de tiempo, sino que llegan dispersas en un rango de tiempo de vuelo. Para minimizar esta dispersión, el tubo de vuelo suele disponer de un dispositivo de enfoque o **reflectrón**, que compensa y corrige esta dispersión de velocidades de los iones de igual razón m/z .

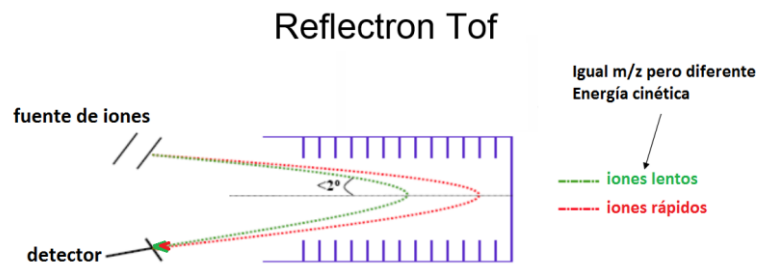


Figura 27. Esquema del funcionamiento de un reflectrón de un espectrómetro de masas de tiempo de vuelo, TOF.

Los iones son dirigidos hacia el reflectrón donde su dirección de vuelo es invertida de manera que un campo de reflexión permite a los iones con mayores energías cinéticas penetrar más profundamente dentro del reflectrón en comparación con aquellos de energías cinéticas menores. Los iones que penetran más en el reflectrón se retardan en su vuelo hacia el detector, con respecto a los que penetran menos. Para un paquete de iones con una misma m/z que contiene iones con diferentes energías cinéticas, el reflectrón disminuirá la dispersión de los tiempos de vuelo de los iones, mejorando la resolución del espectrómetro, en detrimento de la sensibilidad.

Las corrientes de iones que salen del analizador son de una intensidad muy pequeña (del orden de 10^{-8} a 10^{-14} A). Estas bajas intensidades de corriente iónica dificultan la detección, que debe ser rápida y muy precisa, por lo que se requieren detectores muy sensibles como el **detector de impacto electrónico**. Este detector registra la carga inducida o la corriente producida cuando un ión

pasa cerca o golpea una superficie. Consiste en una serie de placas a alto voltaje que convierten el impacto de los iones en señales eléctricas. Como el número de iones que llega al detector en un instante determinado es muy pequeño, es necesario realizar una amplificación significativa para conseguir una señal mínimamente procesable.

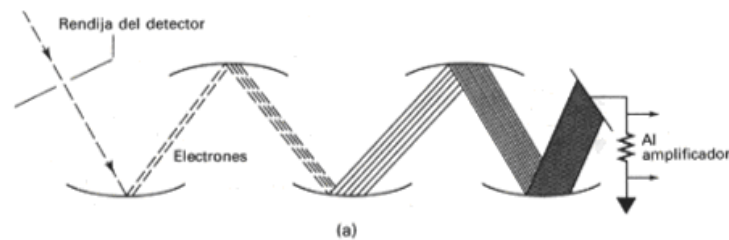


Figura 28. Multiplicador de electrones o electromultiplicador .

La amplificación de la señal generalmente se lleva a cabo mediante un multiplicador de electrones o **electromultiplicador (EM)** que utiliza el principio de emisión secundaria de electrones. Los iones provenientes del analizador de masas son enfocados sobre una placa de óxido de berilio en un sustrato de cobre (Cu-BeO) que emite electrones en proporción directa al número de iones bombardeados hacia él. La corriente de electrones secundarios (electrones generados por impacto de iones) son acelerados y enfocados hacia una segunda placa, la cual vuelve a arrancar electrones secundarios (electrones generados por impacto de electrones) que son acelerados hacia una tercera placa, y así sucesivamente. Generalmente se suelen usar de 10 a 16 placas. De este modo, la amplificación se lleva a cabo a través de un "efecto cascada" de electrones secundarios de placa a placa, donde cada una está conectada a un potencial sucesivamente mayor por un divisor de tensión. Con estos detectores se pueden conseguir amplificaciones de corriente iónica en un factor de 1,000.000 aunque las ganancias típicas son del orden de 100.000.

La salida analógica del detector es una señal analógica de voltaje en función del tiempo de vuelo $V(\text{TOF})$, que mediante un conversor analógico-digital se convierte en una **señal digital**.

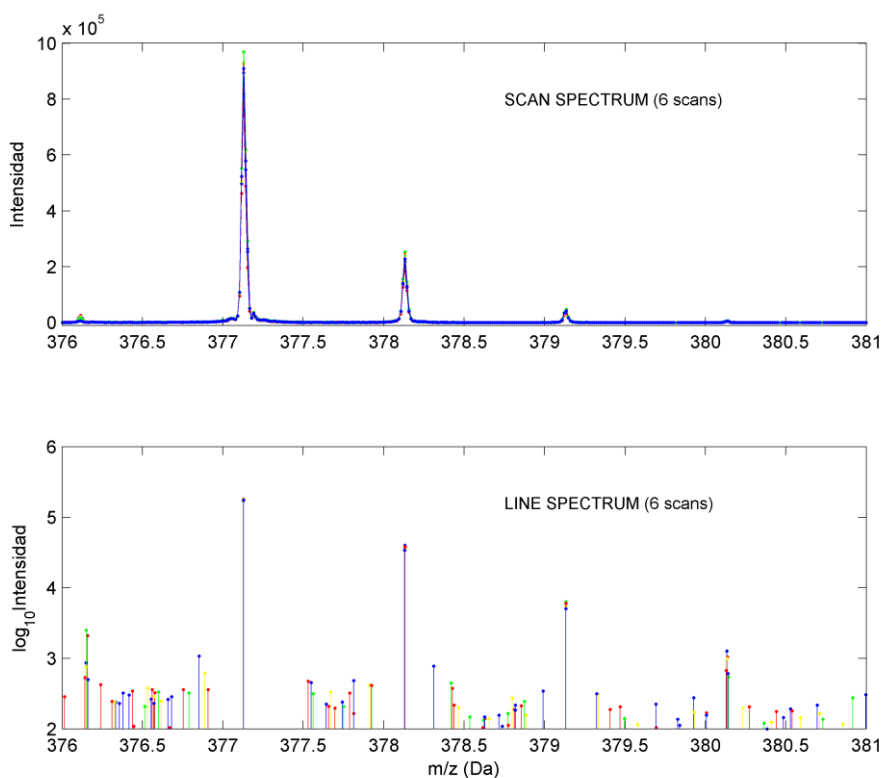
Este tipo de detectores presenta el inconveniente de que la relación entre la señal de entrada (la corriente iónica) y la señal de salida (el voltaje) no es lineal, y para corrientes iónicas elevadas se puede producir saturación de la señal de salida. Gracias a que la velocidad de respuesta del multiplicador de electrones es muy rápida (tiempos de respuesta menores de 50 ns) se pueden realizar miles de **micro-scans** con corrientes de iones pequeñas para evitar la saturación y después obtener lo que se conoce como el **espectro de suma (Sum o Scan spectrum)**. El espectro de suma consiste en una señal discretizada de intensidad o corriente iónica amplificada, convertida a tensión o voltaje $V(\text{TOF})$, calculada como suma de todos valores de $V(\text{TOF})$ de los *micro-scans* para un mismo valor de tiempo de retención.

El número de espectros de suma en un análisis es grande (del orden de miles o decenas de miles). Puesto que los espectrómetros de masas basados en tiempo de vuelo proporcionan precisiones del orden de mDa en m/z , y dado que usualmente se exploran rangos de varios cientos de Da, resulta prohibitivo almacenar las intensidades para todos los posibles valores de TOF o su equivalente la razón m/z . Esto se soluciona haciendo un preprocesado de la señal: se le aplica un filtro paso bajo para eliminar el ruido electrónico y se detectan los valores máximos de cada pico (incluidos picos de ruido) y se aplica un algoritmo que promedia los valores que componen cada pico, para dar los valores de TOF más precisos o su equivalente masa exacta m/z ¹³².

¹³² Urban, J., Kristian, N. & Štys, D., 2014. Trends in Analytical Chemistry fundamental definitions and confusions in mass spectrometry about mass assignment, centroiding and resolution. *Trends in Analytical Chemistry*, 53, pp.126–136.



De este modo, los equipos proporcionan una señal con parejas de valores de intensidad- m/z para aquellas medidas en las que se detectó un máximo en el espectro de suma, lo que limita el número de entradas (en cada tiempo de retención) a unos pocos miles y mejora la sensibilidad del detector. Ésto es lo que se denomina **espectro de línea (*Line Spectrum*)**. Un gran número de entradas presentes en los datos no corresponden a ningún compuesto o fracción del analito, sino que tienen su origen en ruido generado por el instrumento de medida (ruido del detector) o por los procesos mecano-estadísticos a los que están sometidos los iones durante su vuelo (ruido iónico debido a iones que colisionan durante el vuelo e impactan sobre el detector con un TOF anómalo)^{133,134}.



¹³³ Lee, T.A., Headley, L.M. & Hardy, J.K., 1991. Noise reduction of gas chromatography/mass spectrometry data using principal component analysis. *Analytical Chemistry*, (63), pp.357–360.

¹³⁴ Hastings, C. a., Norton, S.M. & Roy, S., 2002. New algorithms for processing and peak detection in liquid chromatography/mass spectrometry data. *Rapid Communications in Mass Spectrometry*, 16(5), pp.462–467.



Figura 29. Arriba: ejemplo de una señal espectro de suma (para 6 *scans* sucesivos), donde se representa en el eje X la m/z en un intervalo de masa (m/z) [375 Da, 381 Da] y en el eje Y la intensidad registrada por el detector para 6 *scans* con $t_R = 1251 \text{ s} - 1256 \text{ s}$, para una muestra de extracto fenólico de aceite de oliva. Abajo: espectros de línea que calcula el espectrómetro a partir de los espectros de suma anteriores. La intensidad está en escala logarítmica.

En la Figura 29 arriba, se ha representado un ejemplo de una señal de espectro de suma, donde en el eje X viene dada la razón m/z y en el eje Y se representa la intensidad registrada por el detector. La señal se ha obtenido como suma de los *micro-scans* en el intervalo de m/z [376 Da, 381 Da], para unos determinados tiempos de retención $t_R = 1251 \text{ s} - 1256 \text{ s}$ (se han representado 6 espectros de suma). Son señales de tiempo discreto y cuantizadas donde se aprecia el ruido que es suma del ruido electrónico y el ruido iónico. En la Figura 29 abajo se representan los espectros de línea que calcula el espectrómetro a partir del espectro de suma, con la intensidad en escala logarítmica. Se observa que en el espectro de línea sólo aparecen los valores de los picos de máxima intensidad correspondientes a un máximo local. Se aprecian claramente cuatro picos que destacan por su gran intensidad, que son relativos a una especie iónica y tres de sus variantes isotópicas. Pero también se registran muchos pequeños picos que probablemente son picos de ruido de la señal o ruido iónico. Se aprecia que el valor de masa exacta que proporciona para cada pico se ha realizado con un promedio de los valores de TOF discretos que se registran en el espectro de suma.

En el análisis con LC-MS, los datos finales registrados consisten en varios miles de espectros de línea que son adquiridos secuencialmente en el tiempo (un espectro de línea para cada tiempo de retención, usualmente, se registra un scan por segundo). Normalmente, se registra un rango de masas de cientos de Da repetitivamente durante un tiempo de análisis cromatográfico. Si por



ejemplo se realiza un examen de muestra cada segundo durante 30 minutos, se obtienen 1800 espectros de línea.

Los datos digitales de espectrometría de masas son **almacenados** codificados como *float* de 32 bits en ficheros específicos de cada fabricante, que pueden ser procesados por los programas que facilitan las casas comerciales de los espectrómetros y también pueden ser convertidos al formato universal XML (*eXtensible Markup Language*). A partir de los ficheros en formato XML (los datos binarios de 32 bits se convierten a formato base 64), se pueden extraer los datos *raw* de los espectros de línea y visualizar los espectros de masas de las muestras analizadas mediante otras herramientas software de tratamiento de señales¹³⁵.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <cmzXML xmlns="http://sashimi.sourceforge.net/schema_revision/mzXML_2.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://sashimi.sourceforge.net/schema_revision/mzXML_2.1 http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/mzXML_idx_2.1.xsd">
- <msRun scanCount="4181">
  <parentFile fileName="E:/UPLC/CIS1.LJ/analysis.bol" fileType="RAWData" fileSha1="c532bf14162271c5f0f4e7c21d10667b1a48f9c2"/>
  <msInstrument>
    <msManufacturer category="msManufacturer" value="Bruker Daltonics" />
    <msModel category="msModel" value="microTOFLC Q" />
    <msIonisation category="msIonisation" value="ESI" />
    <msMassAnalyzer category="msMassAnalyzer" value="Quadrupole TOF" />
    <msDetector category="msDetector" value="MS:1000253" />
    <software type="acquisition" name="micrOTOFcontrol" version="3.0.53.0" />
    <operator first="" last="CIDAF" />
  </msInstrument>
  <dataProcessing centroided="1">
    <software type="conversion" name="CompassXport" version="3.0.5" />
    <software type="processing" name="DataAnalysis" version="4.0.270.0" />
    <nameValue name="recalibrationTime" value="2012-03-28T17:09:29" />
  </dataProcessing>
  <scan num="1" retentionTime="PT1.21100000S" polarity="" msLevel="1" peaksCount="1560" lowMz="48.93918527" highMz="1496.82724056" basePeakMz="232.97507564"
    basePeakIntensity="9073.00000000" tationCurrent="63977.00000000">
    <peaks precision="32" byteOrder="network" pairOrder="m/z"
      int">QkPBuKFwAABCT9t2QaAAEJdfrBKAQAQm8+vUEwAABCFsZdQTAAEKQShBBIAAQpUjRkEwAABClIosQdAAAEKg6T5BcAAAQrPOIKFAAABC1SgUQbAAAEly6j5BIAAAQ900kHAAABC6GQ
    </scan>
  <scan num="2" retentionTime="PT2.21500000S" polarity="" msLevel="1" peaksCount="1543" lowMz="53.00674638" highMz="1502.57777882" basePeakMz="232.97437418"
    basePeakIntensity="8925.00000000" tationCurrent="63824.00000000">
    <peaks precision="32" byteOrder="network" pairOrder="m/z"
      int">QIQG6UGoAABCVkKQJAAAEJrBBBMAAQm2+d0HAAABChr+pQYgAAEKHivjBMAAAQokBaUEwAABCl43rQbgAAEKZh+FBUAAAQptFUHQAAABcnIfqoAAAEKk+sVBIAAAQq/biOGgAABCSw
    </scan>
  <scan num="3" retentionTime="PT3.22000000S" polarity="" msLevel="1" peaksCount="1513" lowMz="46.75707606" highMz="1496.59948185" basePeakMz="232.97509638"
```

Figura 30. Fragmento de un fichero en formato XML conteniendo los datos de espectrometría de masas medidos en el laboratorio.

En la Figura 30 se muestra un ejemplo de un fichero XML asociado a un fichero de análisis LC-MS. Éste presenta una cabecera con información global (fecha de adquisición, configuración del análisis, etc) y a continuación vienen los distintos espectros organizados en bloques para cada tiempo de retención (indicado con

¹³⁵ Shah, A.R. *et al.*, 2010. An efficient data format for mass spectrometry-based proteomics. *Journal of the American Society for Mass Spectrometry*, 21(10), pp.1784–1788.



un índice de esta forma <scan num="1"). En el caso de espectros de línea, para cada tiempo de retención (t_R), el equipo proporciona los datos correspondientes a cada espectro: un índice el tiempo de retención, la polaridad, msLevel, PeaksCount, LowMZ, HighMz, TotIonCurrent,... y los valores de m/z e intensidad.

Si se pone el espectro de líneas en forma de tabla de masas e intensidades (o corrientes iónicas), cada espectro consiste en una tabla con los valores de t_R , m/z , I. Para cada tiempo de retención t_R se obtiene un espectro de masas. En la Figura 31, se muestra un ejemplo de un extracto fenólico de aceite de oliva. Para este análisis, con el método cromatográfico utilizado (cromatografía de partición en fase inversa), aparecen primero los compuestos más polares (t_R bajos) y después los compuestos menos polares (t_R altos). Cada espectro de masas es una lista de valores masa exacta – intensidad (m/z -I) para ese valor de t_R .

t_1 : espectro 1 (muy polar)	$t=1250.39$ N_peaks=6324	$t=1251.39$ N_peaks=6369	$t=1252.40$ N_peaks=6445
t_2 : espectro 2	m= 44.988 Da I= 168.0	m= 44.954 Da I= 188.0	m= 44.929 Da I= 44.0
...	m= 45.091 Da I= 296.0	m= 45.014 Da I= 172.0	m= 44.959 Da I= 52.0
t_N : espectro N (poco polar)	m= 45.136 Da I= 64.0	m= 45.056 Da I= 92.0	m= 45.004 Da I= 96.0
	m= 45.187 Da I= 132.0	m= 45.093 Da I= 112.0	m= 45.035 Da I= 52.0
	m= 45.241 Da I= 64.0	m= 45.164 Da I= 176.0	m= 45.093 Da I= 92.0
	m= 45.297 Da I= 32.0	m= 45.200 Da I= 80.0	m= 45.113 Da I= 172.0
	m= 45.368 Da I= 60.0	m= 45.293 Da I= 220.0	m= 45.199 Da I= 88.0
	m= 45.396 Da I= 80.0	m= 45.350 Da I= 128.0	m= 45.237 Da I= 84.0
	m= 45.405 Da I= 184.0	m= 45.413 Da I= 176.0	m= 45.292 Da I= 32.0
	m= 45.448 Da I= 76.0	m= 45.476 Da I= 92.0	m= 45.335 Da I= 40.0
	m= 45.509 Da I= 160.0	m= 45.547 Da I= 84.0	m= 45.509 Da I= 68.0
	m= 45.556 Da I= 160.0	m= 45.606 Da I= 92.0	m= 45.559 Da I= 28.0
	m= 45.613 Da I= 104.0	m= 45.654 Da I= 80.0	m= 45.599 Da I= 72.0
	m= 45.661 Da I= 100.0	m= 45.698 Da I= 140.0	m= 45.638 Da I= 140.0
	m= 45.770 Da I= 104.0	m= 45.720 Da I= 188.0	m= 45.709 Da I= 192.0
	m= 45.852 Da I= 96.0	m= 45.768 Da I= 164.0	m= 45.743 Da I= 144.0
	m= 45.903 Da I= 52.0	m= 45.823 Da I= 56.0	m= 45.787 Da I= 36.0
	m= 46.021 Da I= 80.0	m= 45.890 Da I= 68.0	m= 45.845 Da I= 72.0
	m= 46.085 Da I= 104.0	m= 45.942 Da I= 284.0	m= 45.928 Da I= 40.0
	m= 46.125 Da I= 168.0	m= 46.051 Da I= 148.0	m= 46.016 Da I= 60.0
	m= 46.160 Da I= 64.0	m= 46.114 Da I= 36.0	m= 46.052 Da I= 128.0
	m= 46.231 Da I= 132.0	m= 46.159 Da I= 272.0	m= 46.098 Da I= 140.0
	m= 46.292 Da I= 64.0	m= 46.197 Da I= 144.0	m= 46.168 Da I= 104.0
	m= 46.354 Da I= 128.0	m= 46.232 Da I= 100.0	m= 46.226 Da I= 60.0
	m= 46.418 Da I= 136.0	m= 46.290 Da I= 188.0	m= 46.258 Da I= 120.0

Figura 31. Ejemplo de datos de espectrometría de masas obtenidos mediante HPLC (fase inversa) acoplada a MS para un extracto polifenólico de aceite de oliva.



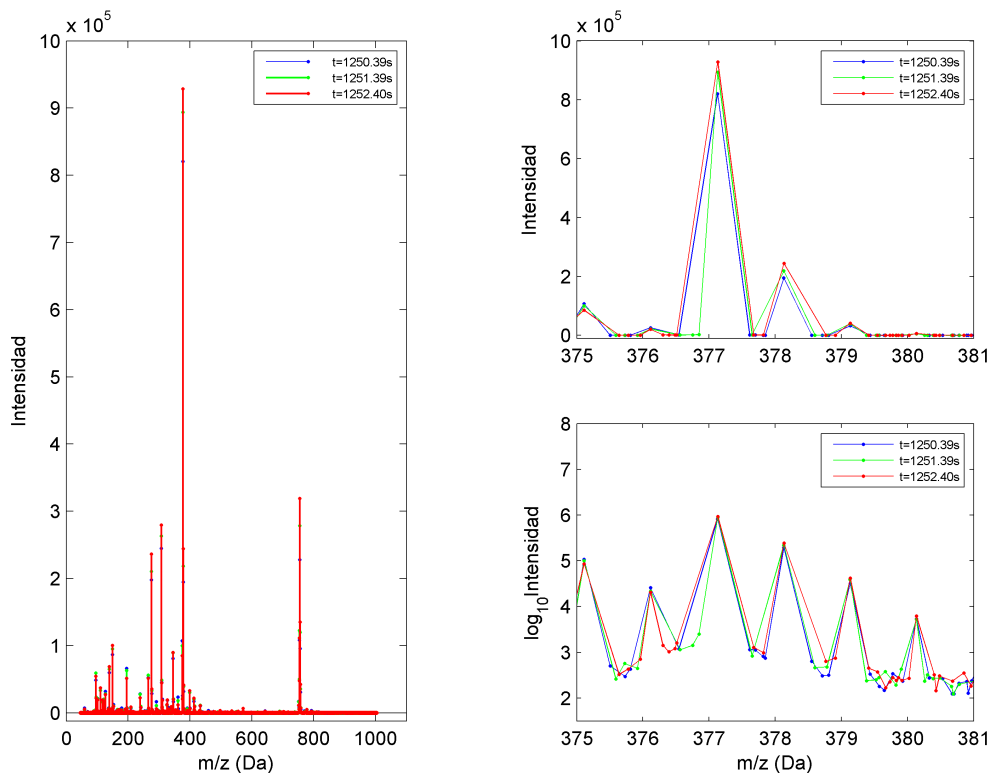


Figura 32. Izquierda: intensidad frente a la masa del ión m/z (Da); en azul un espectro de masas (para 1250.39 s); en verde el siguiente (1251.39 s); en rojo el siguiente (1252.40 s). A la derecha, un detalle para un intervalo de masas pequeño entre 375 y 381 Da. La última gráfica representa la intensidad en escala logarítmica.

En la Figura 32 se muestra un ejemplo de cómo son las señales digitales que forman el *Line Spectrum* de un extracto vegetal analizado por LC-MS. A la izquierda se ha representado la intensidad frente a la masa del ión (relación m/z) de 3 espectros de masas para tres tiempos de retención consecutivos $t_R = 1250$ - 1251 - 1252 s de un extracto fenólico de aceite de oliva. A la derecha se ha representado un intervalo de masas pequeño centrado en $m/z=377$ Da, y la gráfica de la derecha abajo es la misma representación pero en escala logarítmica, donde se aprecian mejor los picos correspondientes a especies iónicas y también el ruido de fondo.

- **Representación de los datos LC-MS.**

Como se ha descrito anteriormente, el cromatógrafo es capaz de separar los compuestos de las muestras según su tiempo de retención (t_R) y el espectrómetro de masas según su relación masa/carga (m/z). En la figura, se representan de forma esquemática los datos crudos (*raw*) obtenidos en el análisis LC-MS: para cada tiempo de retención (t_R) del cromatógrafo se recoge un espectro de masas del espectrómetro (m/z -I).

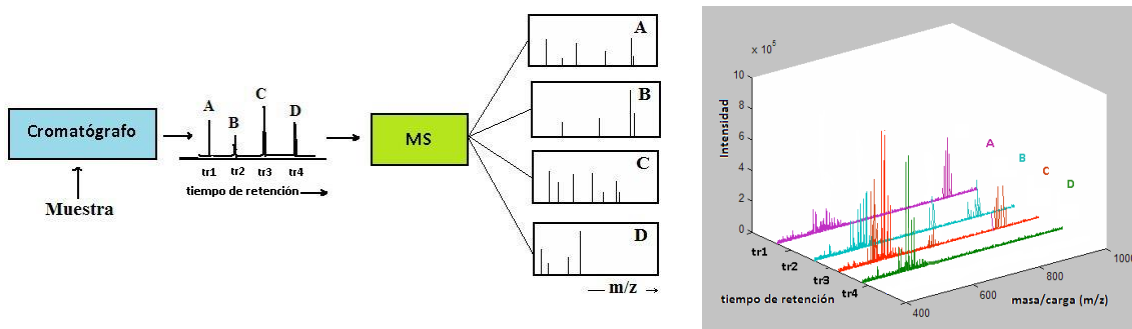


Figura 33. Señales proporcionadas en un análisis LC-MS. A la derecha: cuatro espectros de masas (m/z -I) en diferentes colores para cuatro tiempos de retención determinados (t_R).

La representación típica de los datos de cromatografía acoplada a espectrometría de masas es mediante cromatogramas, que son la representación de la intensidad en función del tiempo de retención cromatográfico.

Cuando se suman las intensidades de todos los iones en cada espectro y se representan como función del tiempo de retención cromatográfico, se le llama **cromatograma de ión total** (*Total Ion Chromatogram*, TIC). Cuando lo que se representa es la intensidad del ión más abundante en función del tiempo de retención se llama **cromatograma de pico base** (*Base Peak Chromatogram*,

BPC). También se puede representar la intensidad a un rango estrecho de m/z , lo que produce un perfil de corrientes de iones seleccionados o **cromatograma de ión extraído** (*Extracted Ion Chromatogram, EIC*). Estas representaciones suelen ser usadas para encontrar componentes en una mezcla compleja sin tener que examinar cada espectro de forma individual.

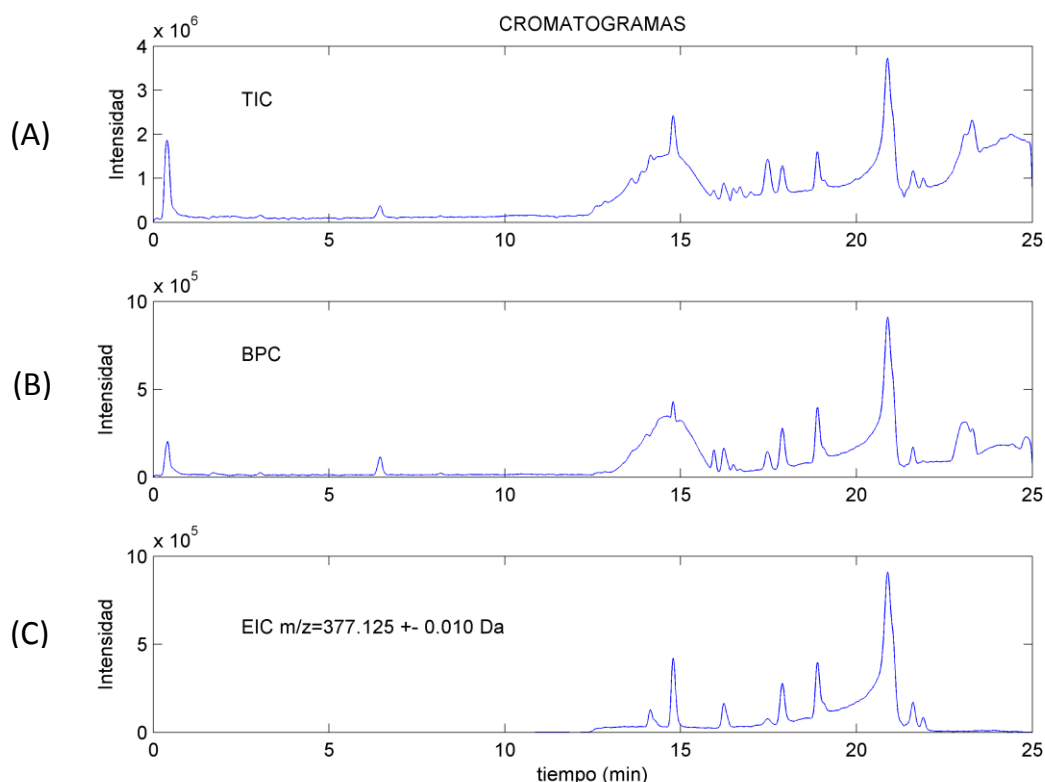


Figura 34. Representación típica de datos LC-MS mediante cromatogramas para un extracto de aceite de oliva. (A) Cromatograma de ión total o TIC: corriente total en función del tiempo de retención para el rango de masas $m/z= 50$ a 1200 Da; (B) Cromatograma de pico base, BPC: corriente correspondiente al ión más abundante en función de t_R ; (C) Cromatograma de ión extraído, EIC: corriente correspondiente a un rango de masas (ión extraído) $m/z= 377.125 \pm 0.010$ Da en función del t_R .

Los datos del espectrómetro de masas se pueden visualizar de forma más global en gráficas 3D mediante lo que se denomina **espectrograma**. Un espectrograma es una representación tiempo de retención-masa-intensidad, donde la

intensidad se muestra mediante un mapa de color. La escala de color representa el logaritmo en base 10 de la intensidad (en este caso, el color rojo se asigna a valores altos y el azul oscuro a valores bajos de la intensidad). Para el cálculo de un espectrograma se elige un intervalo de masa (o resolución de masa) determinado sobre el que se suman las intensidades, y se elige la resolución en tiempo de 1 s, resolución con la que normalmente se programa el espectrómetro para obtener un espectro por segundo.

Los espectrogramas son de alta resolución cuando la resolución en masa elegida es del orden de unos pocos mDa y son de baja resolución cuando se elige un intervalo de masa grande, típicamente de 1Da. En la Figura 35 se representa, a la izquierda un ejemplo de un espectrograma calculado con baja resolución de 1 Da en masa, y 1 s en tiempo de una muestra de un extracto fenólico de aceite de oliva y a la derecha se muestra un detalle sobre ese espectrograma en un intervalo de tiempo de retención [11 min, 23 min] y de masa [370 Da, 384 Da]. A la izquierda de la Figura 36 se muestra un espectrograma de alta resolución con resolución en masa de 50 mDa, y en tiempo de 1 s para la misma muestra. Y a la derecha se muestra un detalle para un intervalo de tiempo de retención y masa.

Esta forma de visualizar los datos mediante espectrogramas, tiene la gran ventaja respecto de los cromatogramas tradicionales de que se pueden ver globalmente todos los datos de una muestra. Ésto nos proporciona información adicional e información relevante de cómo son las señales a tratar y de cómo es su ruido. Con esta representación, de un simple vistazo se pueden identificar las zonas donde se encuentran los iones más abundantes, las zonas de color rojo intenso, y se pueden identificar la ausencia de iones en las zonas de color azul más oscuro.



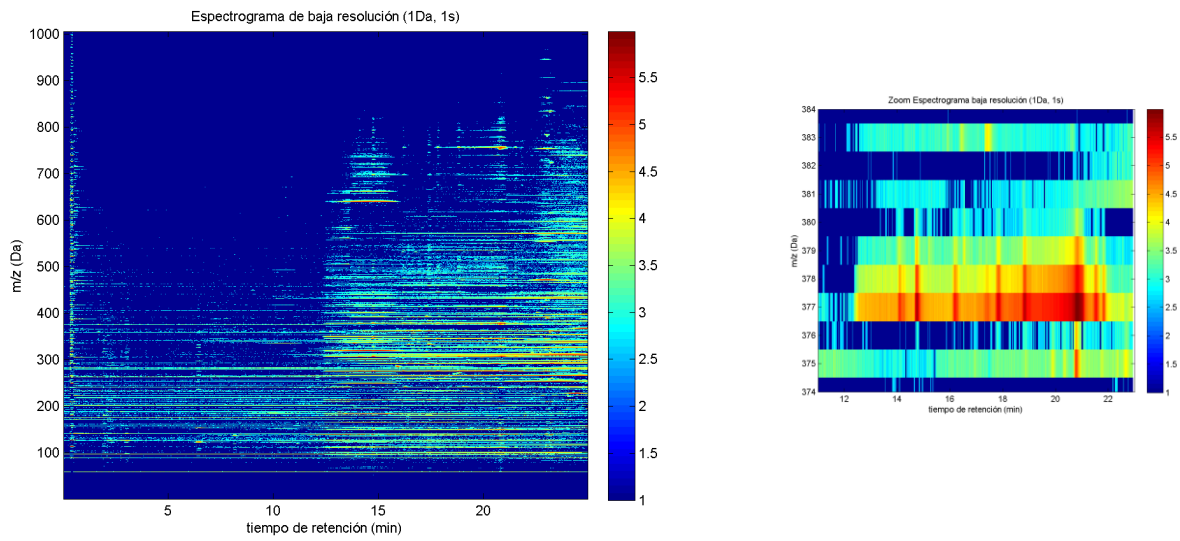


Figura 35. Izquierda: Espectrograma de baja resolución (1 Da, 1 s) para una muestra de un extracto fenólico de aceite de oliva. Derecha: detalle en una zona del espectrograma en el intervalo de tiempo [11 min, 23 min] y de masa [370 Da, 384 Da].

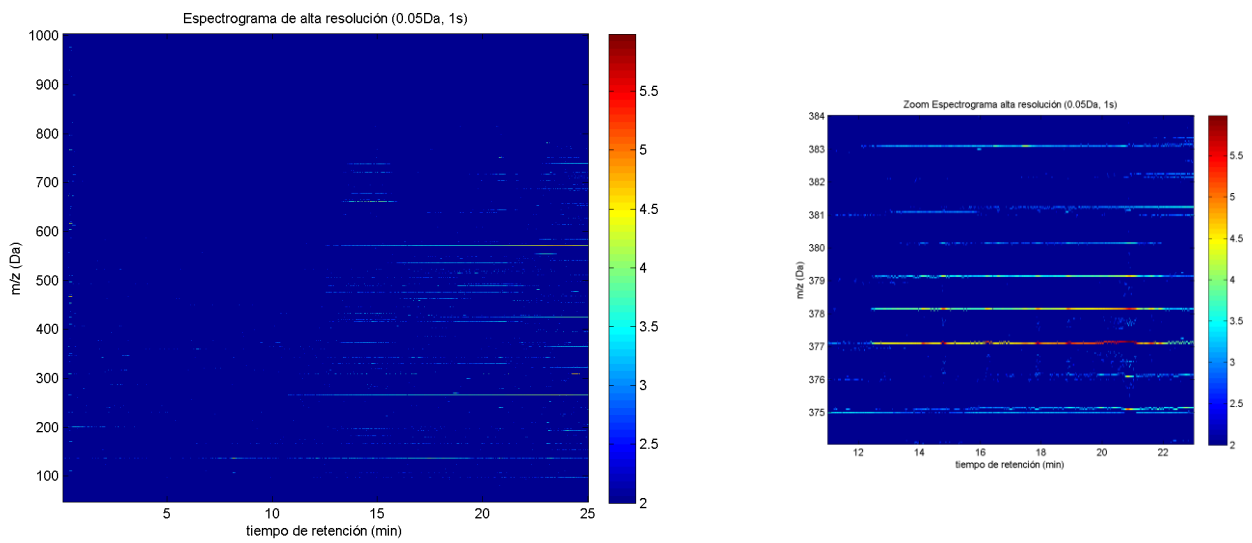


Figura 36. Izquierda: Espectrograma de alta resolución (50 mDa, 1 s) para una muestra de un extracto fenólico de aceite de oliva. Derecha: detalle en una zona del espectrograma en el intervalo de tiempo [11 min, 23 min] y de masa [370 Da, 384 Da].

También se observan muchas zonas con ruido que interfiere y enmascara la señal, son las zonas de poca intensidad de color azul más claro. Hay tantas zonas donde aparentemente el ruido se mezcla con la señal que se hace imposible sin embargo analizarlas de forma visual. El estudio de los compuestos presentes en una muestra se simplifica aplicando técnicas de reducción de ruido.

Para analizar la presencia/ausencia y la identidad de todos los compuestos de una mezcla compleja es necesario en primer lugar eliminar el ruido y después estudiar todo el espectrograma en porciones pequeñas, y analizarlas independientemente. Esto es muy costoso en tiempo y esfuerzo si se hace de forma manual, casi imposible en la mayoría de las ocasiones. Por lo tanto, para estos análisis es necesario recurrir a programas informáticos que sean capaces de recorrer cada zona del espectrograma y analizarla de forma automática separadamente.

- **Características de las señales LC-MS para muestras simples.**

En el caso de muestras formadas por mezclas de pocos compuestos, mediante el análisis LC-MS, es posible asignar unos valores de t_R y m/z concretos a cada pico del cromatograma e identificar los compuestos de forma fácil.

Para tener una visión cualitativa de cómo son las señales LC-MS, se exponen a continuación varios ejemplos de muestras simples; éstas han sido preparadas en el laboratorio utilizando varios estándares comerciales de polifenoles diluidos en metanol para obtener concentraciones determinadas.

El primer ejemplo consiste en una muestra compuesta por un sólo compuesto de masa exacta (m/z) bien conocida, el ácido caféico. En la Figura 37 se muestra arriba el cromatograma BPC (que representa la intensidad del ión más abundante) y abajo el espectrograma de baja resolución (1s, 1Da). Como se aprecia, es sencillo identificar en el cromatograma un único pico y determinar



sus valores experimentales t_R y m/z . La m/z exacta del compuesto se observa a 179.0345 Da, correspondiente al ión $[M-H]^-$ del ácido caféico, ya que el modo de ionización del método es negativo. También se observa en el cromatograma que la señal recogida está contaminada por ruido. Este ruido se aprecia mejor al observar el espectrograma. Se ve que la distribución del ruido no es uniforme, siendo realmente importante para masas bajas, donde aparece como líneas en continuo durante todo el tiempo de análisis. Estas líneas continuas que contaminan el espectrograma y pueden deberse a la fase móvil y la fase estacionaria utilizadas en el proceso cromatográfico. Sin embargo, para masas altas, el ruido puede aparecer como puntos sueltos con una densidad menor asociados a ruido iónico y ruido electrónico.

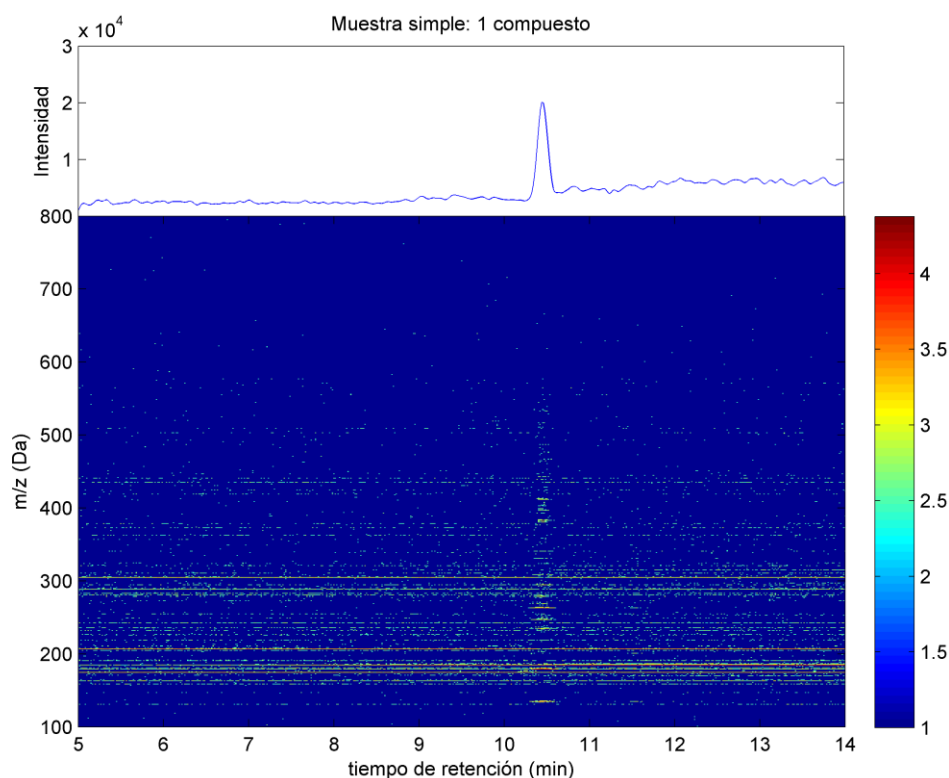


Figura 37. Arriba: cromatograma Base Peak de una muestra simple formada por un solo compuesto, ácido caféico; abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra.



En el espectrograma de la Figura 38, se ha representado un detalle de la figura anterior en torno a $t_R=10.5$ min. Se observa la presencia de una familia de líneas, que están distribuidas, fundamentalmente entre las masas 150 y 500 Da. Estas líneas corresponden al ión principal y a variantes isotópicas, *clusters* y fragmentos del único compuesto que contiene la muestra analizada. Toda esta familia de líneas aparece en un mismo tiempo de retención centrado en $t_R=10.5$ min, durante un intervalo de tiempo de unos 20 segundos.

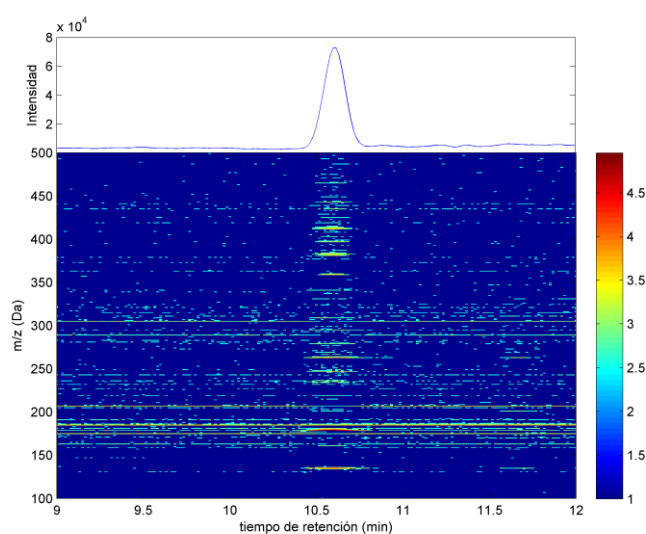


Figura 38. Detalle sobre el espectrograma de baja resolución de la Figura 37 de una muestra simple de 1 compuesto.

El segundo ejemplo consiste en una muestra obtenida mezclando dos estándares, delfinidina-3-O-sambubiósido (con fórmula molecular $C_{26}H_{28}O_{16}$, cuyo ión $[M-H]^-$ tiene masa $m/z=595.1304$ Da) y ácido caféico (con fórmula molecular $C_9H_8O_4$, y cuyo ión $[M-H]^-$, tiene masa $m/z=179.0349$ Da), a distintas concentraciones ($3.0 \mu\text{g/mL}$ y $6.0 \mu\text{g/mL}$, respectivamente). En la Figura 39 arriba se muestra el cromatograma BPC y abajo el espectrograma de baja resolución (1s, 1Da). En este caso, en el cromatograma se ven dos picos muy intensos en los tiempos $t_R=6.2$ min y $t_R=10.5$ min, que se corresponderían a los iones principales de los dos compuestos que componen la muestra y un tercer



pico de menor intensidad en torno al tiempo de retención de $t_R=6.9\text{min}$. Este pico adicional se observa también en el espectrograma en forma de familia de líneas menos intensas. Este pico puede ser debido a la descomposición del compuesto delfinidina-3-O-sambubiósido, o bien a la presencia de un contaminante en la muestra.

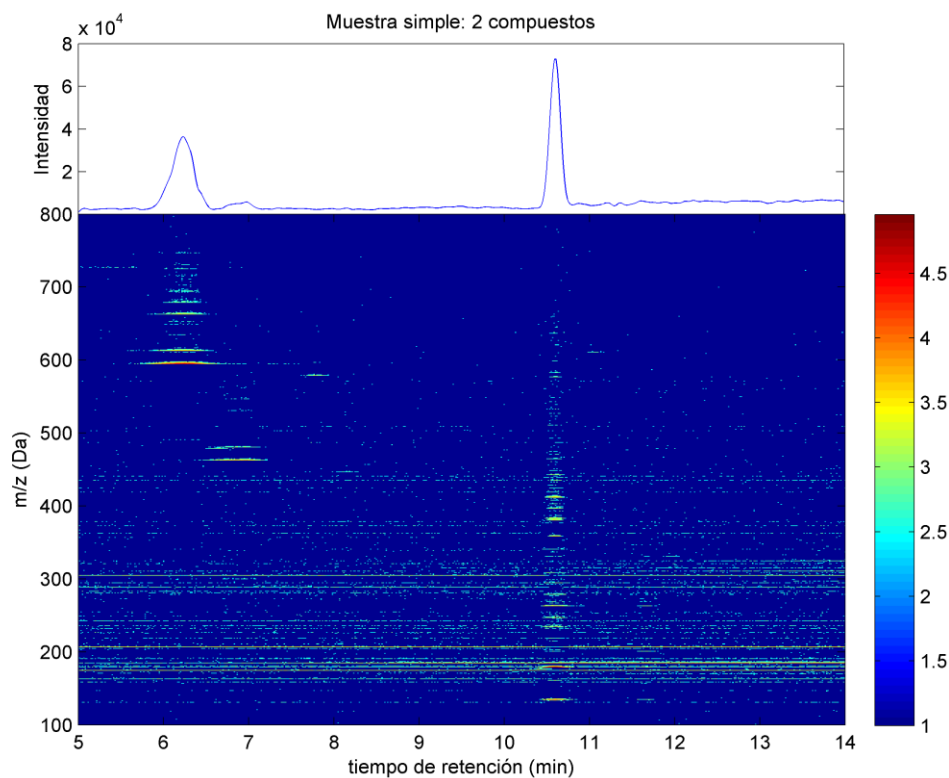


Figura 39. Arriba: cromatograma Base Peak de una muestra simple formada por dos compuestos, delfinidina-3-O-sambubiósido y ácido caféico. Abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra.

Se observa también que la forma de los picos es de tipo gaussiana con diferente anchura para cada uno. El primer pico tiene una desviación estándar en tiempo de unos 10 segundos, mientras que el tercero es más estrecho, tiene una desviación estándar de unos 5 segundos.

Debido a la técnica separativa, los compuestos no llegan al final del elemento separativo en un instante de tiempo, sino que llegan en un intervalo de algunos segundos (dependiendo de las condiciones establecidas y de la movilidad de los compuestos), presentando la concentración, en función del tiempo, un perfil aproximadamente gaussiano o con distribuciones más complicadas. Esta característica hace que durante el intervalo de salida de un cierto compuesto, el espectro de masas presente altas intensidades para los valores de masa correspondientes a los iones, *clusters* y fragmentos asociados a dicho compuesto, y dichas intensidades se deben mantener durante un cierto intervalo de tiempo siguiendo un perfil aproximadamente gaussiano similar a la curva de concentración en función del tiempo de retención proporcionada por la técnica separativa para este compuesto.

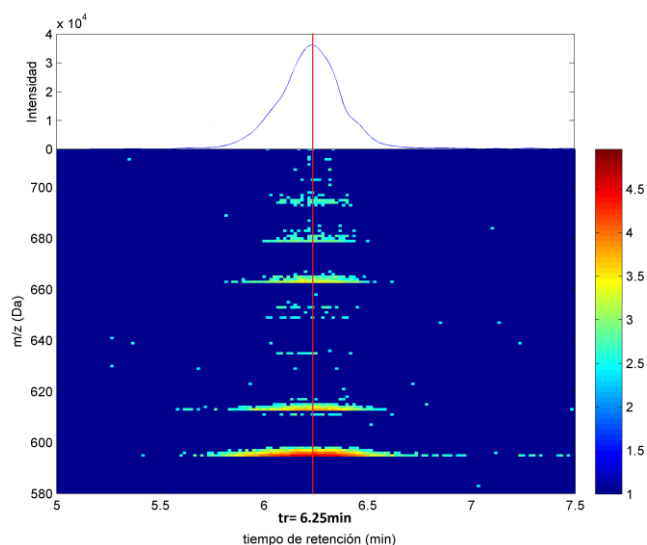


Figura 40. Detalle de la Figura 39 para el intervalo de tiempo [5 min, 7.5 min] y de masa de [580 Da, 710 Da]. Arriba: cromatograma BPC de la delfinidina-3-O-sambubiósido. Abajo, espectrograma de baja resolución (1 s, 1 Da). Se ha marcado el tiempo de retención $t_R=6.25$ min correspondiente al máximo del pico cromatográfico.

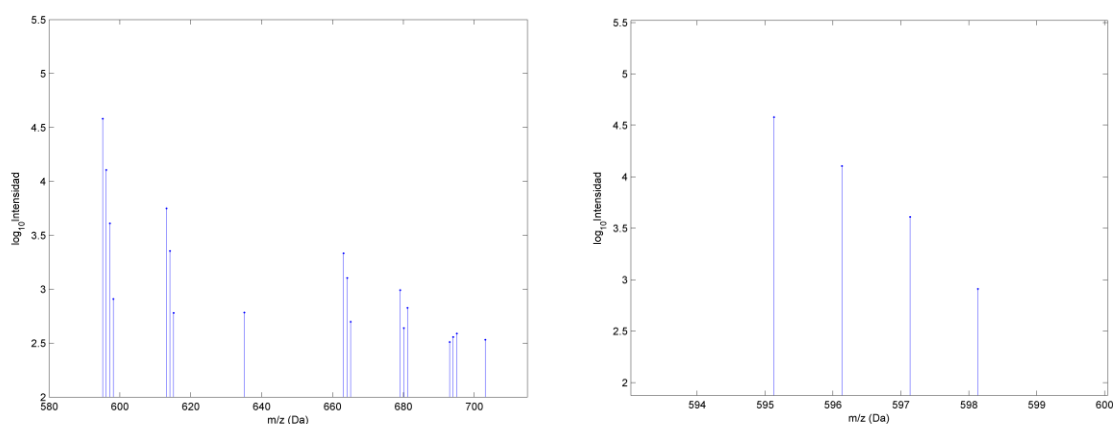


Figura 41. Izquierda: espectros de línea $l(m/z)$ para el tiempo de retención $t_R = 6.25$ min de la Figura 40. Derecha: detalle del espectro de línea de la izquierda, para el ión principal de masa $m/z=595$ Da y sus isótopos $m/z=596, 597, 598$ Da. La intensidad está en escala logarítmica.

En la Figura 40 se representa un detalle de la Figura 39 para el intervalo de tiempo de retención [5, 7.5 min] y de masa [580, 710 Da] correspondiente al pico del ión del compuesto delphinidina-3-O-sambubiósido. En el espectrograma se observa una familia de líneas correspondientes al ión principal, a los *clusters* que se han formado y a sus variantes isotópicas.

En la Figura 41 a la izquierda se han representado los espectros de línea para el tiempo de retención $t_R=6.25$ min, donde se observan varios grupos correspondientes al ión principal y los *clusters*. Dentro de cada uno de estos grupos, se observan las líneas correspondientes a sus variantes isotópicas. Es decir, aparecen los siguientes grupos de líneas: grupo del ión principal y sus variantes isotópicas con $m/z=595-596-597-598$ Da (detalle de la figura de la derecha) y dos grupos de *clusters* con sus variantes isotópicas con $m/z=613-614-615$ Da y $m/z=663-664-665$ Da.

El tercer ejemplo consiste en una muestra obtenida mezclando 5 estándares en metanol, correspondientes a 5 compuestos fenólicos de estándares comerciales



con tiempos de retención y masas bien conocidas a distintas concentraciones. En el cromatograma de la Figura 42 se pueden apreciar claramente 7 picos que se corresponderían 5 de ellos con las 5 moléculas de los estándares que componen la muestra y dos picos adicionales. Estos dos picos adicionales son compuestos desconocidos que pueden deberse a contaminantes de la muestra, a descomposición de una de las moléculas o artefactos cromatográficos. Gracias a que estos picos cromatográficos aparecen resueltos (separados) en el cromatograma es posible estudiarlos fácilmente e identificarlos.

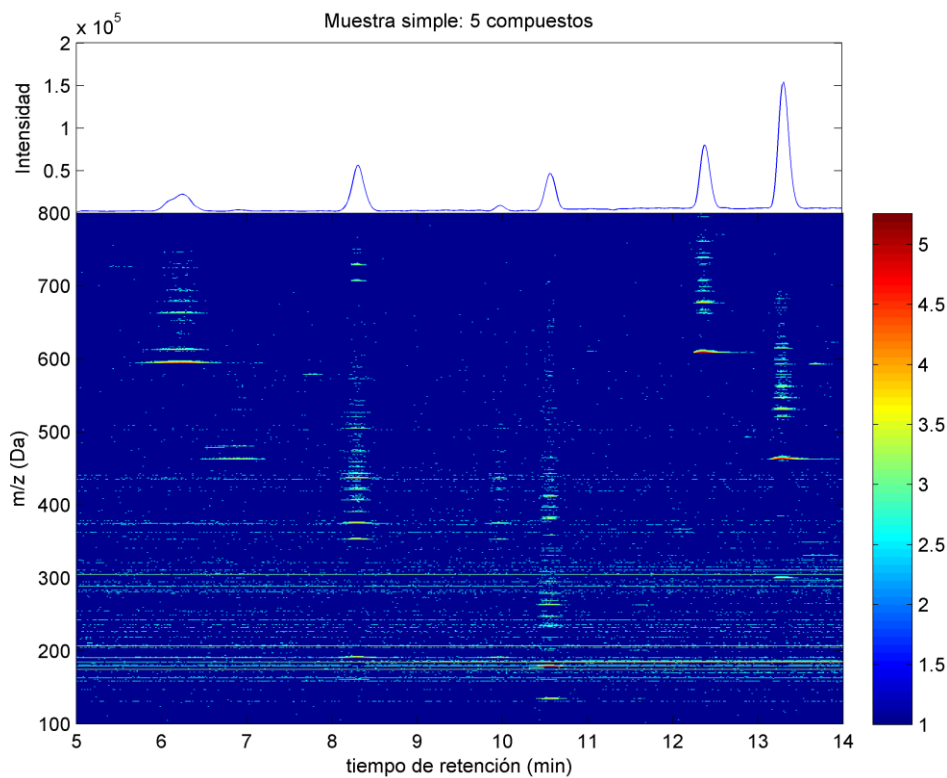


Figura 42. Arriba: cromatograma Base Peak de una muestra simple formada por seis compuestos; abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra simple.

El estudio de los ejemplos de muestras simples con pocos compuestos bien conocidos que ayuda a entender y caracterizar las señales LC-MS. A



continuación, se describen las principales características cromatográficas y espectrométricas de las señales LC-MS.

Características cromatográficas: perfil de pico cromatográfico (eje t_R):

- Idealmente, cada pico del cromatograma está asociado a una molécula de la muestra.
- En la práctica, cada pico del cromatograma puede estar asociado a varias moléculas si no se resuelven los picos (coelución de moléculas).
- Idealmente, cada pico tiene un perfil de pico cromatográfico en forma de gaussiana, cuya anchura depende de las características del proceso cromatográfico y la naturaleza de la molécula. La altura de la gaussiana es dependiente de la concentración y las condiciones de ionización, así como también de la naturaleza de la molécula.
- En los ejemplos estudiados, la desviación estándar del tiempo de retención de los picos cromatográficos es inferior a 20 segundos en todas las moléculas, en el peor de los casos.
- Los perfiles cromatográficos con intensidad casi constante a lo largo de todo el tiempo cromatográfico son artefactos cromatográficos.

Características espectrométricas: huella espectrométrica (eje m/z):

- Una molécula se ioniza idealmente dando un ión, de manera que una línea del espectrograma se correspondería con una corriente iónica con posición correspondiente a la masa exacta del ión.
- En la práctica, en el espectrograma se observan familias de líneas y hay alta probabilidad de fragmentación, formación de dímeros, *clusters*, etc.
- Para cada ión, en función del proceso de ionización y las características de la/s molécula/s presente/s en la cámara de ionización, se observa una familia de líneas espectrales



correspondientes a las variantes isotópicas, de acuerdo con las abundancias relativas de los isótopos de los átomos que forman el ión y la fórmula molecular.

- El conjunto de líneas, para un mismo tiempo de retención y distintos valores de m/z del espectrograma asociadas a una molécula representan **la huella espectrométrica** de la misma.
- Cada línea de la familia de líneas se corresponde con una **especie iónica**.

Una **huella espectrométrica** caracteriza a un compuesto y vendrá descrita por:

- Un perfil cromatográfico común para todas las especies iónicas (t_R común) asociadas a la molécula.
- La masa (m/z) exacta característica de cada una de las especies iónicas que la forman, de acuerdo con la composición de cada ión.
- El área de cada especie iónica dependiente de la concentración y de las condiciones de ionización.
- El área relativa de las variantes isotópicas, dependiente de la abundancia relativa de los isótopos y de la fórmula molecular de cada ión.

Para hacer una representación compacta de la huella espectrométrica, se precisa de la siguiente información: t_R , desviación estándar de t_R , m/z , desviación estándar de m/z , área del ión principal, m/z de los iones menos abundantes y áreas relativas.

El estudio de muestras simples con pocos compuestos (aunque sean desconocidos) es sencillo. Consiste en estudiar cada una de las huellas espectrométricas e identificar las moléculas a través de la identificación de cada especie iónica (m/z exacta y t_R dadas), comparándolas con los valores teóricos de las bases de datos que están específicamente diseñadas para este fin.



Cuanto más complejas son las muestras, los cromatogramas y espectrogramas son mucho más difíciles de interpretar ya que la cantidad de picos a identificar es muy grande y muchos compuestos coeluyen, por lo que son muchos los compuestos que quedan ocultos bajo otros picos de mayor intensidad.

- **Características de las señales LC-MS en muestras complejas de extractos vegetales.**

En la Figura 43 se muestra un ejemplo de una muestra vegetal compleja de un extracto fenólico de pimiento. Se han representado en la parte de arriba los cromatogramas BPC (en rojo) y TIC (en azul) y en la parte de abajo, el espectrograma de baja resolución.

Se observa en los primeros segundos la salida del calibrante, con un pico bien definido en el cromatograma y con una huella espectrométrica o familia de líneas en el espectrograma que abarca un rango de masa muy amplio. En estos cromatogramas, la densidad de picos es muy grande y en el espectrograma se observa mucho ruido de fondo, ruido iónico y electrónico, además de las bandas de ruido que aparecen de manera continua durante todo el tiempo de análisis. (artefactos que se deben a la fase móvil que arrastra la muestra a través de la columna separativa durante todo el proceso analítico). En el espectrograma también se observan huellas espectrométricas o familias de líneas correspondientes a posibles compuestos, representadas como líneas más intensas (con tonalidades que van del amarillo al rojo).

Si se hace un análisis convencional, se consideraría manualmente cada uno de los picos mayoritarios del cromatograma BPC y se identificarían como posibles compuestos de la muestra. Con esta forma de trabajar, se conseguiría identificar del orden de 50-100 compuestos, en el mejor de los casos. La utilización del espectrograma permite identificar de forma supervisada muchas más especies



iónicas. Sin embargo, lo más probable es que no sean detectados muchos posibles compuestos debido a estar enmascarados por el ruido. Por otra parte, un estudio supervisado sería un trabajo muy costoso.

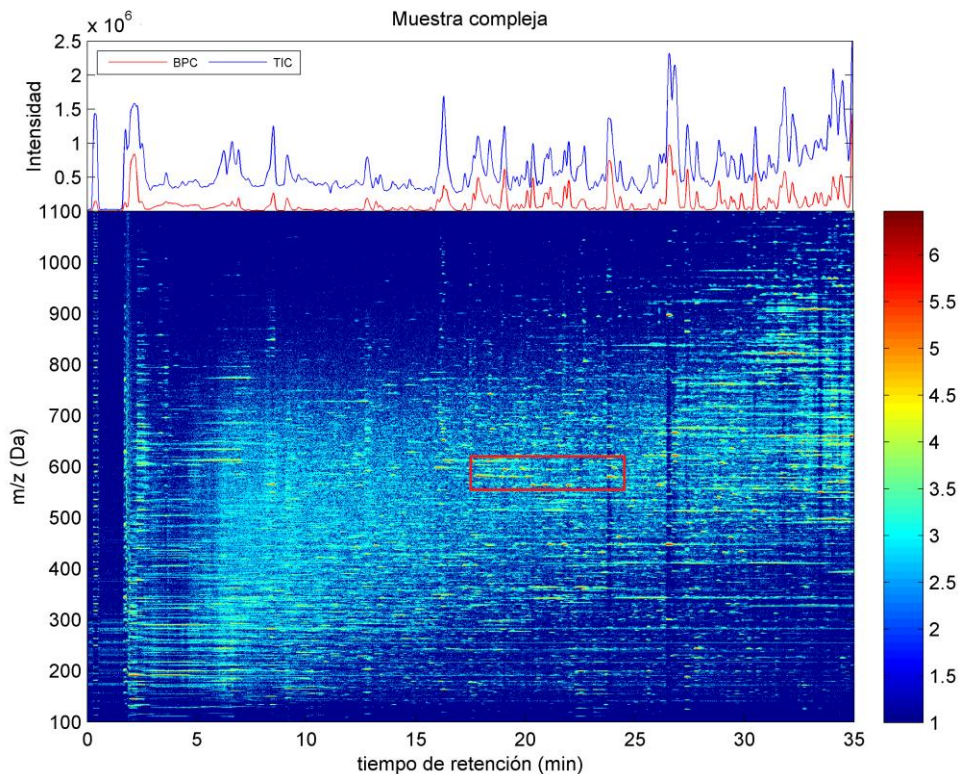


Figura 43. Arriba: cromatogramas BPC (en rojo) y TIC (en verde) del análisis LC-MS de un extracto fenólico complejo de pimienta. Abajo: espectrograma de baja resolución de dicho extracto vegetal. El cuadro marcado en rojo corresponde a un detalle que se muestra en la figura siguiente.

En la Figura 44, se ha representado un detalle de la Figura 43 en una zona comprendida por el intervalo de tiempo [17 min, 26min] y un rango de masas m/z [560 Da, 630 Da]. Se han representado arriba los cromatogramas de pico base (BPC) y de ión extraído (EIC) para una masa $m/z=597.1\pm 0.1$ Da y abajo el espectrograma de baja resolución. Sólo en ese cuadrado, que representa una pequeña porción de todo el espectrograma, se pueden identificar de forma

visual más de 100 especies iónicas con intensidad elevada (en color rojo y naranja). Un efecto de la respuesta del espectrómetro de masas, debido a la incertidumbre del tiempo de vuelo de los iones es que la m/z exacta queda bien definida para picos de mucha intensidad, pero los picos de poca intensidad se ensanchan mucho y no es fácil encontrar una ventana en masa (m/z) adecuada para obtener el cromatograma de ión extraído. Este ensanchamiento se aprecia en el ejemplo de la Figura 45 donde se ha representado el cromatograma EIC para la masa $m/z=447.1\pm 0.1\text{Da}$ y su representación m/z en función de t_R .

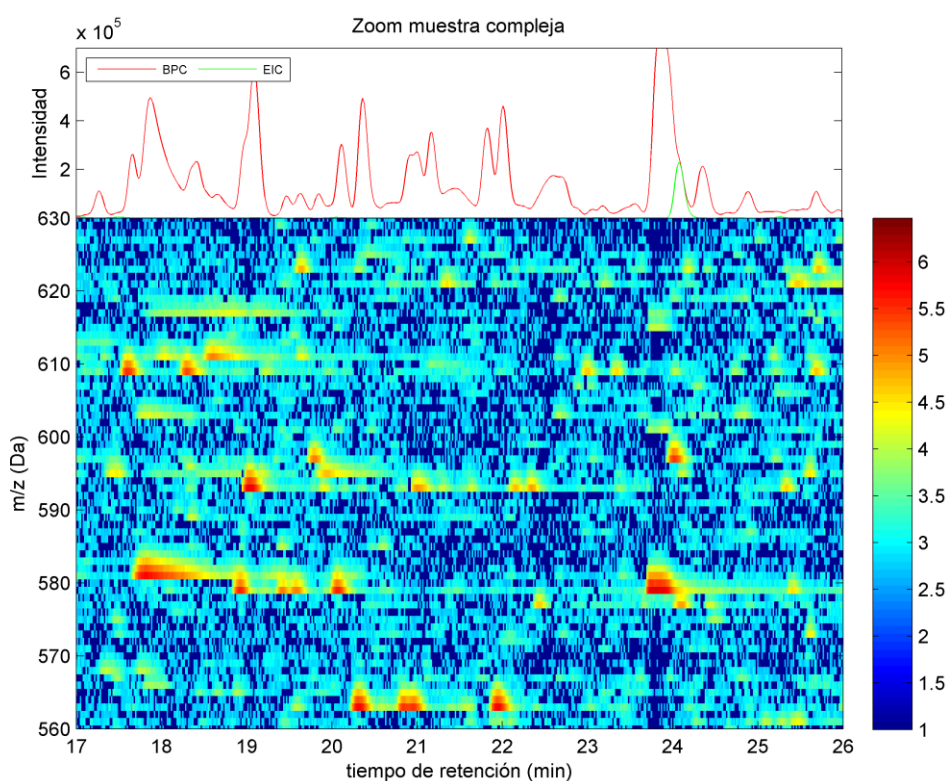


Figura 44. Arriba: cromatogramas; BPC (en rojo) de una muestra vegetal compleja de un extracto vegetal de pimienta y cromatograma EIC (en verde) para una masa $m/z=597.1\pm 0.1\text{ Da}$.

Abajo: espectrograma de baja resolución de la misma muestra vegetal.

Si se quisiera analizar cada una de las especies iónicas por separado manualmente se debería de acometer la tarea de hacer del orden de 10^4

cromatogramas de ión extraído con ventanas de masa de 5mDa para este intervalo de m/z y del orden de 10^5 para todo el espectrograma completo. Este tipo de análisis resulta prohibitivo y la solución requiere diseñar nuevas herramientas para procesar los datos de forma automática.

En los cromatogramas de la Figura 44 se observa que en el cromatograma EIC para la masa $m/z=597.1\pm 0.1\text{Da}$ hay un pico (línea en verde) para el tiempo de retención $t_R=24\text{min}$. Este pico aparece oculto debajo de un gran pico del cromatograma BPC (línea en rojo) que representa la intensidad del ión más abundante. Por lo tanto en el análisis convencional estudiaría el pico grande en rojo a $t_R=24\text{min}$, pero estaría fuera de nuestro alcance visual el pico oculto debajo.

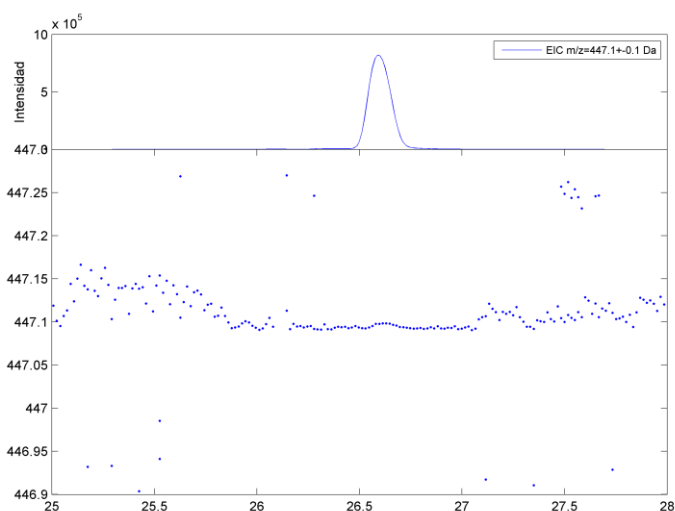


Figura 45. Arriba: cromatograma EIC para $m/z=447.1\pm 0.1\text{ Da}$. Abajo: representación de la masa m/z en función del tiempo de retención, donde se aprecia que m/z exacta queda bien definida en la zona del pico donde tiene mucha intensidad.

Las zonas de color oscuro del espectrograma de la Figura 44 corresponderían con gaps o zonas libre de ruido de de datos. Dichas zonas son más fácilmente apreciables en espectrogramas de alta resolución. Si se consigue eliminar parte



del ruido de fondo, los gaps serán mayores y se presentaran tanto en el eje de tiempo de retención (t_R) como en el de masa (m/z), dejando aisladas las regiones que contienen las distintas especies iónicas. Estos gaps se pueden aprovechar para localizar las distintas especies iónicas y así poder analizarlas separadamente.

De la observación de los espectrogramas y cromatogramas representados para los ejemplos analizados, se sabe que un pico (asociado a una especie iónica) de una determinada relación m/z aparece en un tiempo t_R dado como una nube de puntos (entradas del fichero t_R - m/z) agrupados que presentan una desviación estándar de masa del orden de $dm/z=5$ mDa (la resolución del espectrómetro de masas) y una desviación estándar de tiempo de retención de unos $dt_R=10$ segundos aproximadamente. Estas observaciones se han aprovechado para diseñar los algoritmos que detectan picos/compuestos o ruido.

E.2) Nuevas herramientas de tratamiento de datos LC-MS: métodos de *clustering*.

La situación ideal para el estudio de una muestra es que la separación cromatográfica sea completa, es decir que en cualquier instante sólo eluya, como mucho, una molécula. En este caso, cada molécula tiene un tiempo de retención distinto del resto, en el cromatograma aparecen picos bien resueltos y la identificación de los compuestos se puede hacer de forma convencional: para cada pico con un valor de t_R dado se estudia su espectro de masas que proporciona la huella espectrométrica del compuesto, y al ser única para cada compuesto, es fácil su identificación.

En muestras complejas, la técnica cromatográfica no es capaz de separar todos los compuestos y para un tiempo t_R dado, pueden eluir simultáneamente varias moléculas. La caracterización convencional por picos será ciega a ciertos



compuestos que coeluyen, ya que se encuentran ocultos bajo otros picos de mayor intensidad. Por tanto, el método convencional de detección de compuestos no es una herramienta adecuada para la identificación de compuestos en muestras complejas debido al problema de la coelución. Además, este problema se agrava por el ruido.

El **objetivo** del tratamiento de los datos LC-MS se puede plantear en los siguientes términos:

*"dados unos ficheros con los datos crudos (raw) resultado del análisis LC-MS, obtener una **lista todos los compuestos** con la huella espectrométrica de cada uno de ellos, en la presencia de ruido y artefactos"*

Para conseguir este objetivo es necesario diseñar nuevas herramientas que realicen un tratamiento automático de los datos de LC-MS y que tengan como objetivos:

- Eliminar el ruido, procurando no perder información importante de las especies iónicas.
- Considerar las características cromatográficas y espectrométricas de las especies iónicas descritas anteriormente.
- Detectar los picos cromatográficos sin romper especies iónicas.
- Detectar los picos ocultos que coeluyen.
- Tener en cuenta los gaps o zonas de ausencia de datos en t_R y m/z .
- Tener en cuenta la información estructural de los compuestos: posibilidad de variantes isotópicas, fragmentos, *clusters*, dímeros, etc.
- Realizar un tratamiento automático y no muy costoso en tiempo computacional, ya que en muchos estudios es necesario estudiar una gran cantidad de muestras.



- Proporcionar herramientas estadísticas y de visualización de los resultados.

El algoritmo desarrollado para el tratamiento de los datos LC-MS sigue el esquema que se muestra en la Figura 46. Para ello, se utiliza el entorno matemático MatLab, que es ampliamente usado para el tratamiento de señales de todo tipo. Con esta herramienta se puede desarrollar cualquier programa que sea necesario para un estudio específico.

El primer objetivo antes de realizar el tratamiento de los datos LC-MS es realizar un **preprocesamiento** consistente en tres pasos: primero, extraer los datos crudos, transformar los ficheros del espectrómetro de masas a XML y éstos, a formato binario para facilitar su manipulación con MatLab, eliminando información no relevante y reduciendo el tamaño; segundo, filtrar la señal para la eliminación del ruido y tercero, calibrar la masa (m/z).

Los datos preprocesados se ordenan en una matriz de datos: t_R , m/z , I y su tratamiento consiste en distinguir los picos correspondientes a posibles compuestos de otros picos del ruido de fondo, lo que se llama **clustering, peak picking o peak detection**^{114,116,117,125}.



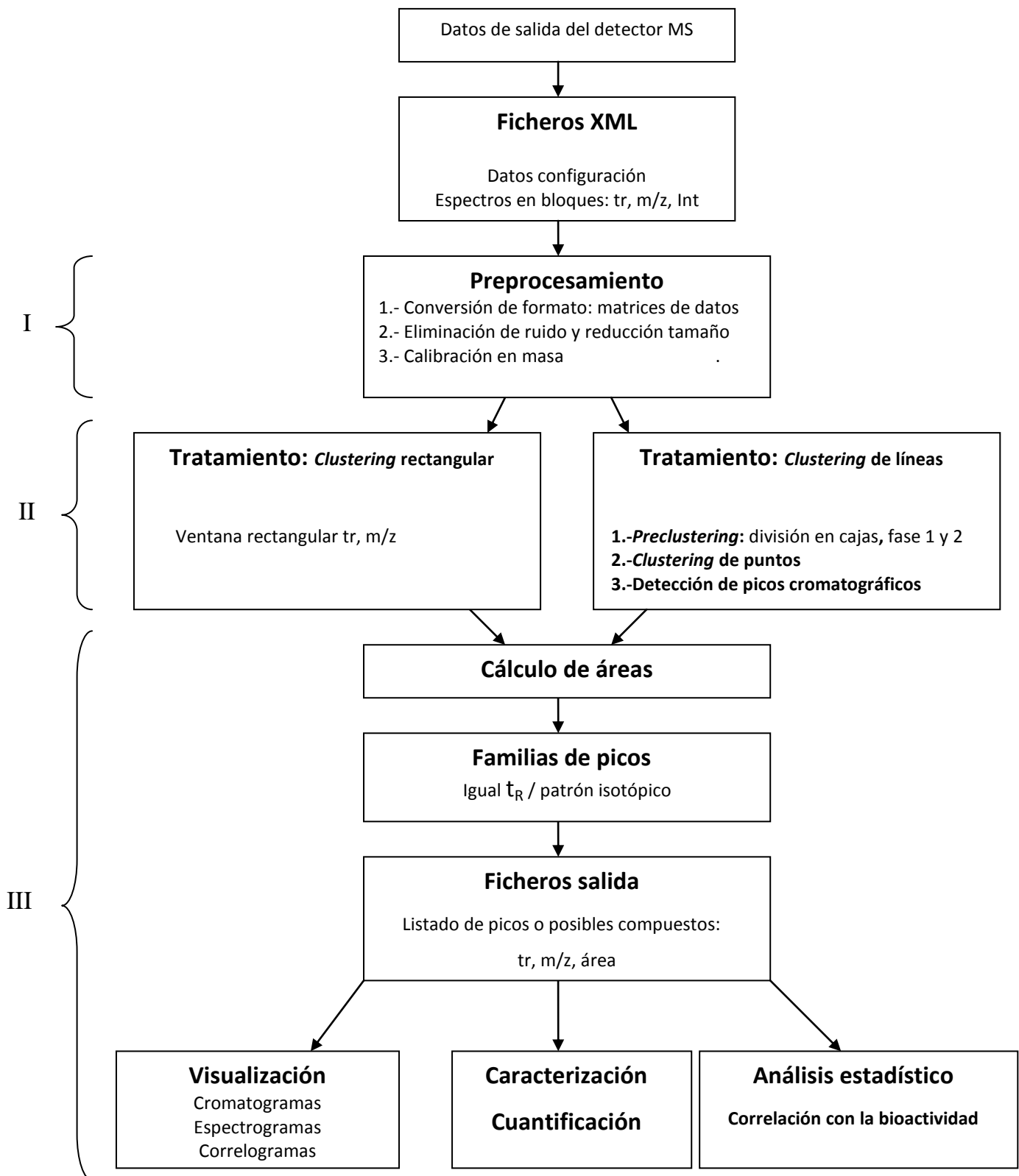


Figura 46. Esquema seguido para el tratamiento de datos LC-MS desarrollado: métodos de *clustering* rectangular y *clustering* de líneas.



En esta tesis doctoral, se propone un nuevo método de tratamiento para la detección de picos de LC-MS, que se basa en un concepto muy simple, "divide y vencerás". Consiste en hacer porciones del espectro de m/z vs. t_R , separándolo en cajas de forma recursiva hasta llegar a las unidades mínimas y evaluar las distancias entre las cajas. Las cajas estarán definidas por una anchura de valor de tiempo y una altura de valor de masa (m/z). Las cajas tienen que ser clases disjuntas en el espectro de MS, es decir, cada elemento pertenecerá a una sola caja. Y después, se analizan los picos cromatográficos dentro de cada caja. Este método es un método recursivo, automático y muy rápido computacionalmente.

De forma integrada con los programas de tratamiento de las señales se calcula el área cromatográfica de cada uno de los picos y se utiliza la información de la regla del patrón isotópico para poder agrupar en familias los picos detectados con igual tiempo de retención y una diferencia de m/z de 1,004 Da en promedio o 0,502 Da en dobles ionizaciones. Esto facilita las comparaciones entre distintas muestras complejas, puesto que basta con elegir el compuesto de mayor abundancia de una familia para realizar las comparaciones, lo que ahorra mucho tiempo computacional.

El último paso del tratamiento de las señales consiste en generar una lista de *clusters* o posibles compuestos, donde se especifican los valores: t_R , m/z exacta, con sus desviaciones estándar y su área. A partir de esta lista, se pueden representar los datos de diferentes formas, caracterizar o cuantificar la muestra, realizar estudios estadísticos o comparar distintas muestras.

Como se observa en el esquema de trabajo descrito (Figura 46), a partir de los datos crudos LC-MS, se llevan a cabo la estrategia propuesta, que puede dividirse en tres grandes etapas: preprocesamiento (I), tratamiento (II) y



procesamiento (III). A continuación, se detalla cada uno de los pasos que tienen lugar dentro de cada etapa.

I. Preprocesamiento.

Preprocesamiento 1. Conversión de los datos a formato binario.

Para poder trabajar con los datos es necesario ordenarlos en listas de vectores o matrices y en formato binario. Por ello, se extraen los datos LC-MS crudos (con extensión .baf) y se transforman a XML mediante un programa proporcionado por el fabricante del espectrómetro de masas. Se ha creado un programa que extrae del fichero XML los datos de cada espectro (índice, valores de configuración del equipo, número de líneas detectadas, instante de tiempo de medida y conjunto de pares de valores m/z -intensidades para ese tiempo de medida t_R) y a continuación, se almacenan los tripletes tiempo-masa-intensidad (t_R - m/z -I) en un fichero con formato binario (con extensión .bin).

Preprocesamiento 2. Eliminación de ruido.

Los ficheros binarios obtenidos ocupan decenas de MB antes de quitar ruido, por lo que por medio de esta conversión. Se aplica un procedimiento de detección de picos correspondientes a ruido y eliminación de éstos, que reduce significativamente el tamaño de los ficheros quedando reducido el problema del almacenamiento, transporte y envío de los datos^{117,119,123}.

Las señales de m/z y tiempo de retención (t_R) registradas por los espectrómetros de masas tienen mucho ruido, como se ha descrito anteriormente. El algoritmo propuesto para la eliminación de ruido explora, para cada tiempo-masa-intensidad, la presencia de puntos en una ventana de tiempo (de unos pocos segundos) y una ventana de masa (de unos pocos mDa). Un punto se considera válido únicamente si se han observado un número mínimo de puntos en el



intervalo de tiempo y masa establecido. Si no aparecen varios puntos en un intervalo de t_R y m/z próximos, no se trata de una especie iónica, puesto que no cumple con las características esperadas de acuerdo con el modelo de generación de puntos ($t_R, m/z, I$) establecido. Estos puntos son máximos aislados correspondientes a ruido electrónico e iónico y, por lo tanto, se eliminan del fichero de datos¹³⁶⁻¹³⁸.

El tamaño de las ventanas de tiempo se establece teniendo en cuenta la técnica separativa y el periodo de análisis del conjunto de muestras (que determina la anchura típica de los picos en el tiempo) y la ventana de masa, en función de las características del instrumento de medida de tiempo de vuelo.

Adicionalmente, se puede complementar el criterio de eliminación de ruido con un umbral de intensidad. Se realiza una eliminación de puntos no relevantes, seleccionando los puntos en los que el número de cuentas (intensidad) supera un umbral establecido. A partir del histograma de intensidad de una muestra, se puede establecer un umbral de intensidad adecuado a cada muestra estudiada. En la Figura 47 se ha representado el histograma de la intensidad de un extracto fenólico de pimiento analizado por LC-MS. En este ejemplo, se aprecia que la frecuencia de intensidad es muy elevada para valores inferiores a 100 cuentas. Seleccionando este umbral, se elimina el ruido de fondo y además queda garantizado que no se pierde ningún pico de interés, ya que los picos de interés tienen valores de intensidad muy superiores a 100 cuentas.

¹³⁶ Danielsson, R., Bylund, D. & Markides, K.E., 2002. Matched filtering with background suppression for improved quality of base peak chromatograms and mass spectra in liquid chromatography–mass spectrometry. *Analytica Chimica Acta*, 454(2), pp.167–184.

¹³⁷ Smith, C.A. *et al.*, 2006. XCMS : Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, 78(3), pp.779–787.

¹³⁸ Li, Y., Qu, H. & Cheng, Y., 2008. An entropy-based method for noise reduction of liquid chromatography–mass spectrometry data. *Analytica Chimica Acta*, 2, pp.19–22.



De esta forma, fijando un umbral de intensidad bajo, se puede hacer una supresión de ruido más conservadora (preserva entradas de datos que presenta intensidad baja) y fijando un umbral de intensidad alto se hace una supresión de ruido más agresiva (sólo preserva entradas de datos de mayor intensidad). El umbral adecuado se debe establecer en función del estudio a realizar con los datos y de los objetivos de dicho estudio.

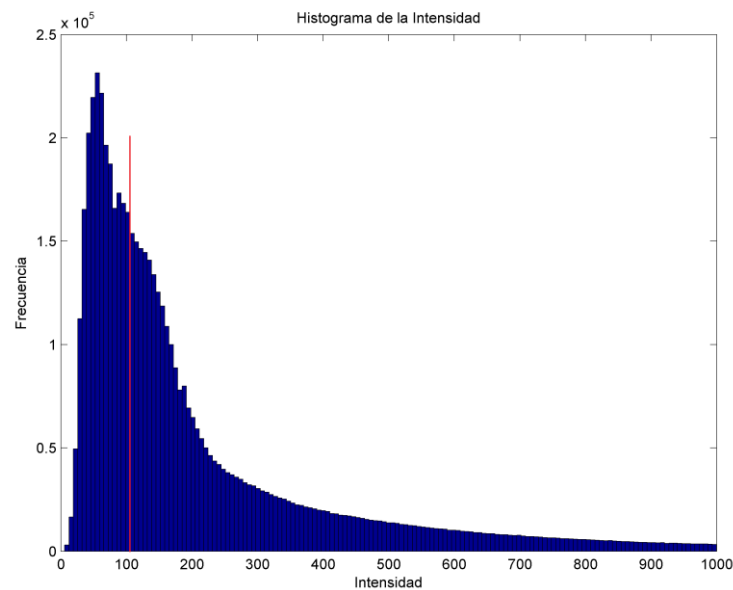


Figura 47. Histograma de la intensidad de una muestra de un extracto vegetal de pimienta.

Como ejemplo de aplicación del algoritmo de reducción de ruido propuesto, se ha aplicado sobre un espectro de masas obtenido del extracto fenólico de pimienta. Se ha tomado una ventana temporal de 3 s (que, para una frecuencia de muestreo de un espectro por segundo, se traduce en el análisis del espectro actual, el anterior y el siguiente). La ventana de masa se ha establecido en 100 mDa. Una entrada se ha considerado válida si en los espectros anterior y posterior se ha observado una entrada con masa igual a la de referencia ± 50 mDa. Adicionalmente, para estudiar el efecto del umbral de intensidad, se han aplicado umbrales desde 0 cuentas (no se aplica umbral) hasta 1000 cuentas.



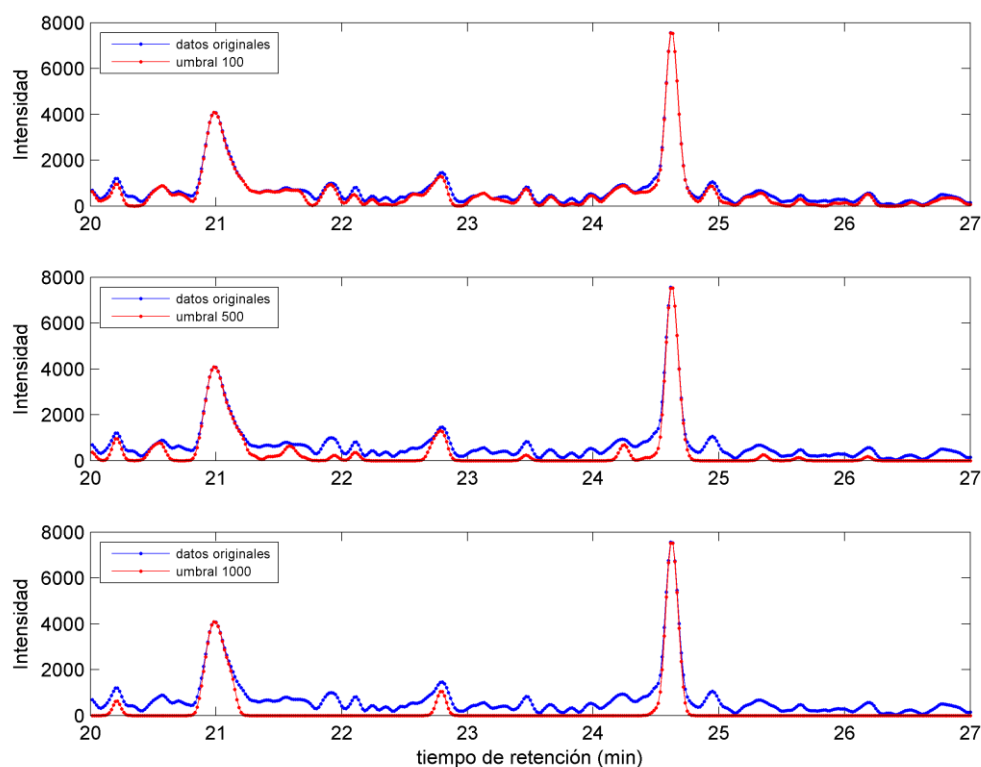


Figura 48. Cromatogramas EIC para $m/z = 449 \pm 0.5$ Da de un extracto vegetal de pimienta, aplicando el algoritmo de reducción de ruido con distintos umbrales de intensidad, 100, 500, 1000 cuentas.

En la Figura 48, se muestran los cromatogramas de EIC para $m/z = 449 \pm 0.5$ resultantes de aplicar el algoritmo de reducción de ruido (en azul), comparados con el cromatograma de EIC original (en rojo) para distintos valores del umbral de intensidad aplicado. Puede observarse que, para un umbral conservador (100 cuentas), el ruido de fondo se reduce de forma sensible, y que, incluso con un umbral más agresivo (1000 cuentas) los principales picos del cromatograma permanecen inalterados.

El efecto del algoritmo de reducción de ruido puede apreciarse mejor observando el espectrograma de alta resolución. La Figura 49 muestra el espectrograma de alta resolución (1s, 10mDa) correspondiente al intervalo

centrado en la masa $m/z=449$ Da. En las distintas figuras se aprecia de arriba a abajo: los datos originales y la reducción de ruido aplicando distintos umbrales de intensidad: 50, 100, 200, 500 y 1000 cuentas. Se observa como el algoritmo elimina progresivamente las entradas menos significativas, preservando las entradas del espectro de masas más relevantes. Los espectrogramas de baja resolución de la Figura 50, con resolución de (1s, 1Da) sobre el rango 200 Da a 700 Da, ilustran igualmente como la aplicación del algoritmo elimina entradas correspondientes a ruido, preservando la información relevante del espectro de masas. Dependiendo de cada estudio, conviene elegir un procedimiento de eliminación de ruido más o menos agresivo. Por ejemplo, para la caracterización de un extracto vegetal, es preferible hacerlo sobre un espectrograma con reducción de ruido 1000, pero para la identificación de compuestos bioactivos es preferible usar una reducción de 100 para asegurar que no se pierden posibles candidatos minoritarios.



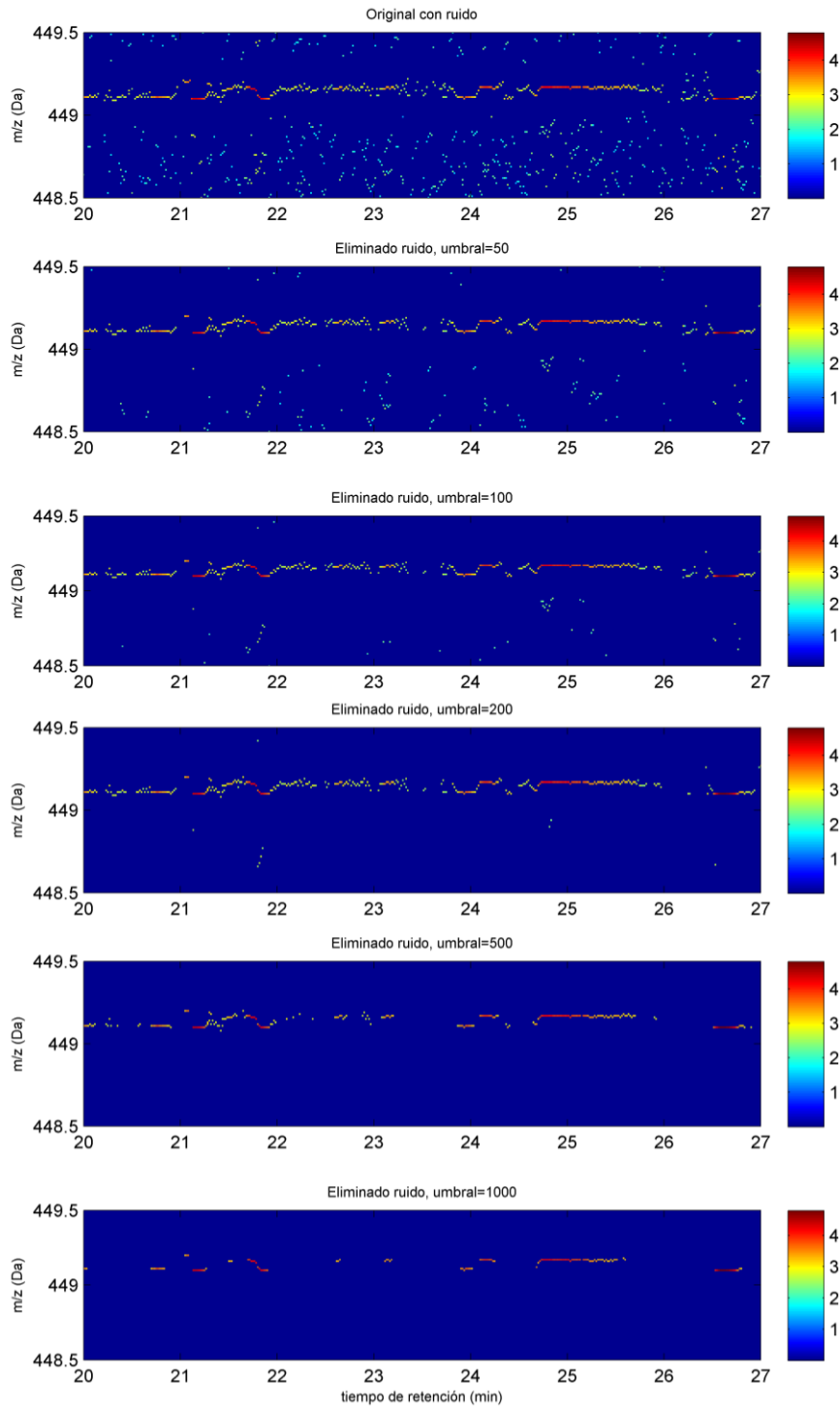


Figura 49. Espectrogramas de alta resolución (1 s, 10 mDa) para la masa $m/z=449$ Da de un extracto vegetal de pimienta analizado por LC-MS. Se ha aplicado el algoritmo de reducción de ruido con distintos umbrales de intensidad: 0 (no se quita ruido), 50, 100, 200, 500 y 1000 cuentas. La escala de color representa $\log_{10}(\text{intensidad})$.

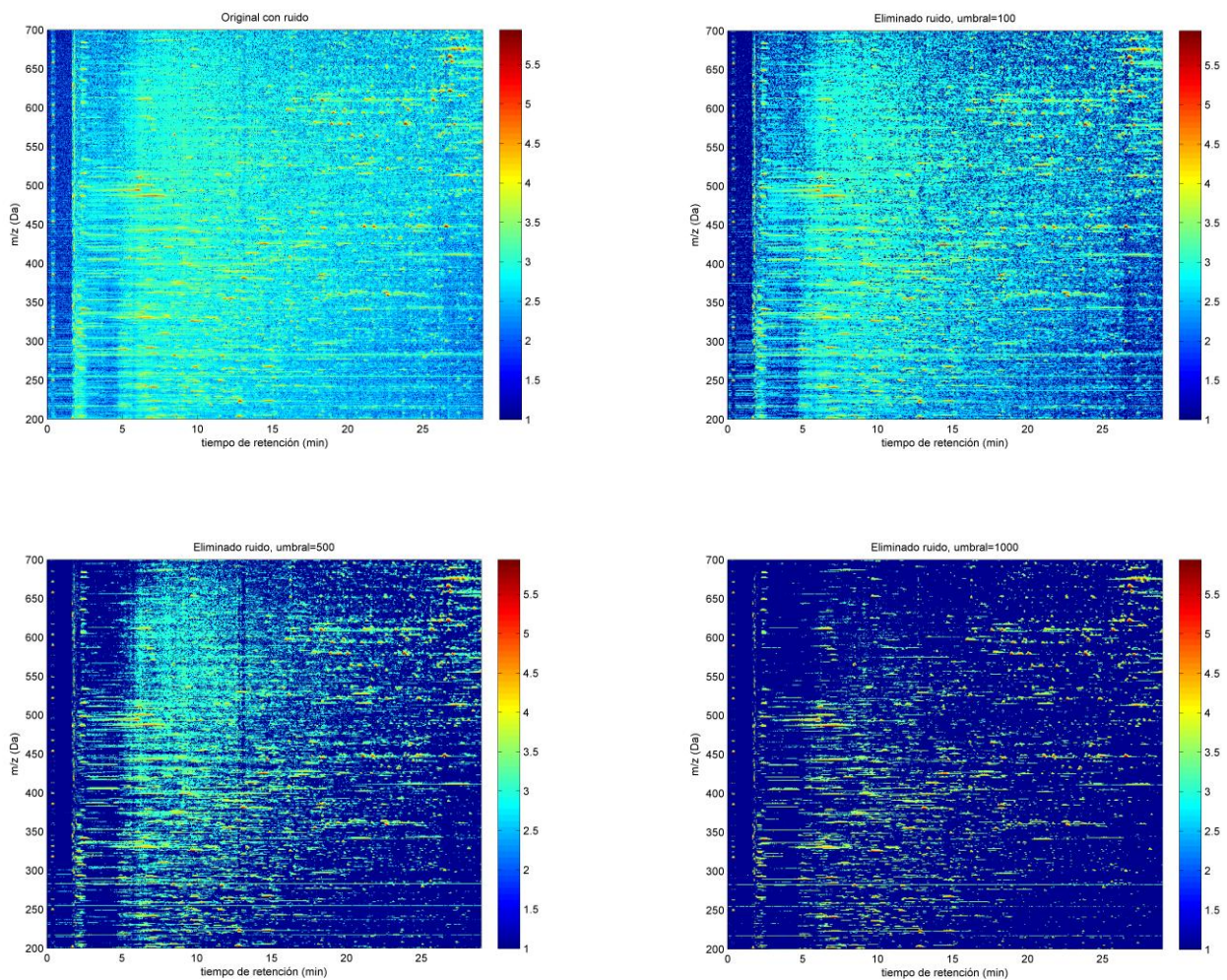


Figura 50. Espectrogramas de baja resolución (1 s, 1 Da) para el rango de masas [200 Da –700 Da], aplicando el algoritmo de reducción de ruido con distintos umbrales (de izquierda a derecha y de arriba abajo), espectrograma original (con umbral 0), y espectrogramas con umbral 100, 500 y 1000. Corresponde a un extracto fenólico de pimienta.

Como resultado de la aplicación del algoritmo de ruido, en función del umbral aplicado, el número de entradas presentes en los ficheros de espectrometría de masas se ve reducido considerablemente. La Tabla 4 muestra el volumen de datos que queda tras aplicar el algoritmo de reducción de ruido en los ficheros de espectrometría de masas de la muestra de pimienta. Se observa que el algoritmo (sin aplicar umbral) es capaz de eliminar un 65% de los datos, todos

ellos correspondientes a ruido con una alta probabilidad y que, a medida que se aplica un umbral mayor, el volumen de datos decrece considerablemente.

Tabla 4. Efecto del algoritmo de reducción de ruido sobre el tamaño de un fichero binario (.bin) de un extracto vegetal de pimiento. El fichero binario se ha obtenido a partir de un fichero .baf del LC-MS que ocupa 553.29 MB.

Reducción ruido	Umbral intensidad	Tamaño fichero	Número puntos ($t_R, m/z, I$)	Número puntos normalizado
No	0	136.80 MB	17.505.213	100.00%
Sí	0	47.41 MB	6.062.332	34.66%
Sí	50	38.34 MB	4.902.257	28.03%
Sí	100	21.64 MB	2.763.799	15.82%
Sí	500	4.48 MB	567.899	3.27%
Sí	1000	1.82 MB	227.292	1.33%

Preprocesamiento 3. Calibración de la masa.

Para poder realizar la calibración de los valores de masa-carga (m/z) de los registros del LC-MS, normalmente se introduce en el análisis de la muestra un calibrante o sustancia de referencia, cuyo espectro de masas es bastante complejo, pero muy bien conocido. El calibrante abarca el rango de masas que se quiere analizar y aparece en los primeros segundos de los registros. Comparando los valores experimentales y teóricos de m/z del calibrante, se obtiene el ajuste lineal o polinómico necesario para calibrar las m/z de todos los espectros de masas registrados.

II. Tratamiento de los datos LC-MS: métodos de *clustering*.

Una vez realizado el preprocesamiento de los datos de espectrometría de masas, se dispone de un conjunto de ficheros de datos ($t_R, m/z, I$) almacenados en forma matricial y donde se han eliminado entradas correspondientes a ruido,



con un tamaño mucho más reducido, por lo que son más fáciles de manejar. Se procede después al tratamiento de los datos con el objetivo de determinar cada uno de los compuestos que componen la muestra e identificar cuál o cuáles de esos compuestos es o son los responsables de una bioactividad determinada mediante el análisis estadístico apropiado.

A continuación, se exponen dos métodos de *clustering* que se han desarrollado, en base a las características de las señales LC-MS estudiadas. El primero de ellos, el *clustering* rectangular, se basa en los espectrogramas de baja resolución. Es un *clustering* sencillo que no hace consideraciones sobre las características de la señal. El otro método, el *clustering* de líneas se centra en la idea fundamental de transformar el espectrograma en un conjunto de **clusters** (o zonas donde hay datos con intensidad elevada y no son ruido) de manera que se cumpla que en cada *cluster* puede haber una o varias especies iónicas, pero una especie iónica no puede pertenecer a dos *clusters*.

- **Método de *clustering* rectangular.**

Este método de *clustering* es una manera sencilla y rápida de obtener *clusters*. Se basa en dividir los datos LC-MS mediante ventanas de tiempo y masa rectangulares en cajas que formarán los *clusters* y que vendrán definidos por un valor de tiempo y masa dado por el valor central de la caja y un valor de área para el conjunto de los puntos que se encuentran dentro de la caja. Por ejemplo, se puede considerar una ventana de varios segundos de resolución en tiempo de retención y 1 Da de resolución en m/z .

En este método, se ha estimado el área de la intensidad asociada a cada *cluster* $(t_i, m/z_j)$ como la integral de la intensidad sobre un intervalo de tiempo y masa m/z centrado en $(t_i, m/z_j)$. Para una muestra n , se estimó el área $A_n(t_i, m/z_j)$



aplicando para m/z una ventana rectangular de ± 0.5 Da alrededor de m/z_j y para el tiempo de retención, una ventana Hamming de ± 7.5 s alrededor de t_i :

$$A_n(t_i, m/z_j) = \int_{\tau=-7.5}^{+7.5} h(\tau) \int_{\mu=-0.5}^{+0.5} I_n(t_i + \tau, m/z_j + \mu) d\tau d\mu \quad (\text{Ecuación 4})$$

donde $h(\tau)$ es una ventana Hamming de 15 s. Se utiliza la ventana de Hamming para suavizar los picos cromatográficos y compensar las derivas del tiempo de retención entre distintas muestras de cara al estudio de colecciones de muestras.

Este método realmente lo que hace es transformar el espectrograma en un conjunto de rectángulos o *clusters* de igual tamaño, donde cada rectángulo viene representado por un valor de área. Con ello, se consigue acotar y reducir el problema del tratamiento de los datos LC-MS.

Este método de *clustering* está orientado a la detección de bioactividad ya que se pueden hacer estudios estadísticos de correlación entre las áreas de los *clusters* y los datos de bioactividad. Sin embargo, no es adecuado para la identificación de especies iónicas o compuestos porque aunque se reduce el número de datos a tratar, en algunos casos sigue siendo muy elevado.

Como ejemplo de este método de *clustering* rectangular, se ha calculado el área asociada a los *clusters* rectangulares obtenidos con una ventana de anchura 7 s (con desplazamiento de 1 s) y altura 1Da a una muestra simple compuesta por 5 estándares comerciales de polifenoles (Figura 51). Se ha representado el área en escala logarítmica con una escala de color. Cada *cluster* o rectángulo viene representado por un valor del área. En la Figura 52 se muestra un detalle en el que se observan los *clusters* rectangulares con su color correspondiente a un valor de área determinado.



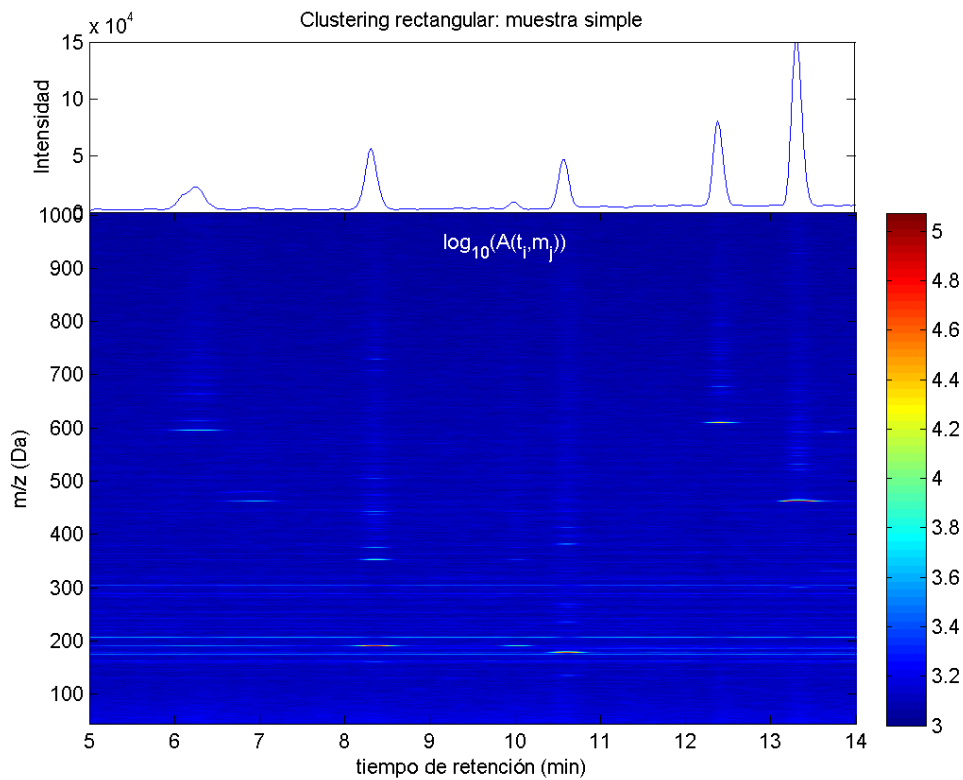


Figura 51. *Clustering* rectangular aplicado a una muestra simple formada por 5 estándares comerciales de polifenoles. Arriba: cromatograma BPC. Abajo: representación del área de los *clusters* rectangulares, calculada con una ventana rectangular de anchura $t_R=1$ s y altura $m/z=1$ Da. El área viene representada con una escala de color. La escala representa el logaritmo en base 10 del área.

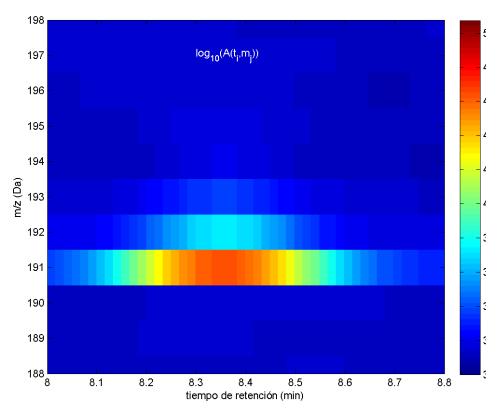


Figura 52. Detalle del *clustering* rectangular de la Figura 51 aplicado a una muestra simple. Se representa el área de los *clusters* rectangulares, con anchura de 1 s y altura de 1 Da. El área viene representada con una escala de color.



- **Método de *clustering* de líneas.**

El segundo método de *clustering* que se propone para la detección de compuestos de los datos de LC-MS se denomina *clustering* de líneas. Este método de *clustering* se basa en dividir el conjunto de líneas o puntos $(t_R, m/z, I)$ del LC-MS en porciones más pequeñas, que son más fáciles de analizar separadamente, hasta llegar a las unidades mínimas o *clusters* que se corresponderán con las posibles especies iónicas. En este proceso de división, se realiza un *preclustering* que, partiendo del conjunto de datos completo, divide el espectro de forma recursiva en cajas en el eje de tiempo y en el eje de m/z , en base a la presencia de gaps o intervalos de t_R o m/z , donde no se observan líneas. Después, se relacionan las cajas en función de la medida de la distancia entre ellas. Este *preclustering* es muy robusto y no tiene riesgo de cometer errores, ya que no se elimina información: sólo consiste en reagrupar los puntos $(t_R, m/z, I)$ de manera que cada punto pertenezca a una sola caja o *cluster*.

La gran ventaja de este *preclustering* basado en cajas es que una vez estructurados los datos en cajas, el análisis no trabaja con muchas líneas sino con muchas cajas que contienen pocas líneas, lo que hace el cálculo menos costoso computacionalmente, y por otra parte, el nivel de complejidad es controlable: si las cajas son grandes, el proceso es rápido y poco detallado. Si las cajas son pequeñas, el proceso es lento.

Cuando se llega a unidades mínimas indivisibles mediante cajas, se recurre a la división de picos cromatográficos mediante la determinación de máximos y mínimos locales de la función $I(t_R)$, intensidad vs. t_R , de cada caja. Este análisis tiene un riesgo implícito de pérdida de información y es un cálculo mucho más lento computacionalmente. Pero el hecho de realizar la detección de picos cromatográficos para cada caja aisladamente (obtenida del *preclustering*) hace



que el riesgo no sea tan grande como cuando se realiza el método de detección de picos sobre el cromatograma, como se hace en el método convencional.

Por lo tanto, la novedad y la aportación más importante que se hace al tratamiento de los datos de LC-MS es el *preclustering* y la obtención de una lista final de *clusters* o posibles compuestos con sus valores ($t_R, m/z, I$) que facilita considerablemente la labor de detectar compuestos. Entre los compuestos del listado, se encuentran los mayoritarios, todos los minoritarios (de baja intensidad) y, además, el método es capaz de detectar los picos ocultos bajo otros de mayor intensidad, los cuales no son detectables con el método convencional.

A continuación, se describe detalladamente el método desarrollado de *clustering* de líneas. Primero, se describe el *preclustering* que incluye varios procesos de división del espectro de masas (división en cajas), después el *clustering* de puntos y por último, se describe la técnica para la detección de picos cromatográficos.

Paso 1 del *clustering* de líneas. *Preclustering* basado en cajas.

A partir del conjunto de datos ($t_R, m/z, I$) de LC-MS se forman cajas: con una resolución en tiempo de retención dada (T_{resol}), se buscan los gaps o ausencia de masas m/z por donde partir y formar las cajas. Se forman así cajas descritas por los siguientes parámetros: esquinas de la caja (valores, $t_{R_{\text{min}}}, t_{R_{\text{max}}}, m/z_{\text{min}}, m/z_{\text{max}}$) y los puntos ($t_R, m/z, I$) que están dentro de cada caja. Se considera la relación entre cajas mediante el criterio de distancias entre ellas, aplicando los umbrales de diferencia en tiempo (dt_R) y en masa (dm/z) más adecuados para cada estudio. Según dicho criterio, las cajas que se toquen (se acerquen) entre sí, pertenecerán a un mismo *cluster* y las que no, definirán un nuevo *cluster*. Los *clusters* son disjuntos, es decir, cada punto ($t_R, m/z, I$) pertenecerá a un solo



cluster. A cada *cluster* se le asigna un índice y unos parámetros que lo identifican del resto.

Se consigue así clasificar las cajas y agrupar los datos (t_R , m/z , I) en *clusters*, garantizando que queden en el mismo *cluster* todos los puntos (t_R - m/z - I) asociados a uno o varios picos cromatográficos. Hasta aquí, es un proceso sin riesgo, no se eliminan datos y cada división es segura (no hay riesgo de que una especie iónica esté en dos *clusters*) si los gaps son grandes. Después de esta primera fase, el tratamiento posterior de muchos *clusters* pequeños es más fácil de resolver.

En la práctica, es preferible realizar un primer análisis de *clustering* de cajas con ventanas temporales grandes. Se trata de lo se denomina "*preclustering*: fase 1": con $T_{\text{resol}}=300$ s, y una relación de distancia entre cajas con umbrales $dt_R=20$ s y $dm/z=20$ mDa. De esta manera, se empieza con pocas cajas grandes y el cálculo es más rápido. En un segundo análisis, se realiza un *preclustering* con una ventana menor, "*preclustering*: fase 2": con $T_{\text{resol}}=20$ s con los mismos umbrales de distancia entre cajas, $dt_R=20$ s y $dm/z=20$ mDa.

Como ejemplo de aplicación del *preclustering*, se aplica a una muestra de extracto polifenólico de pimiento. En la Figura 53 a la derecha se muestra el espectrograma de baja resolución (1 s, 1 Da) con reducción de ruido y con un umbral de intensidad de 100 cuentas. Se ha marcado la zona a estudiar con un rectángulo rojo de anchura 15min (900s) y altura 4 Da. La figura de la derecha es el espectrograma de alta resolución (1s, 10mDa) de la zona que vamos a estudiar. Se observan zonas de alta densidad de puntos formando líneas en torno a 447, 448, 449 Da, correspondientes a compuestos, y grandes gaps o zonas de ausencia de datos en el eje m/z entre las líneas.



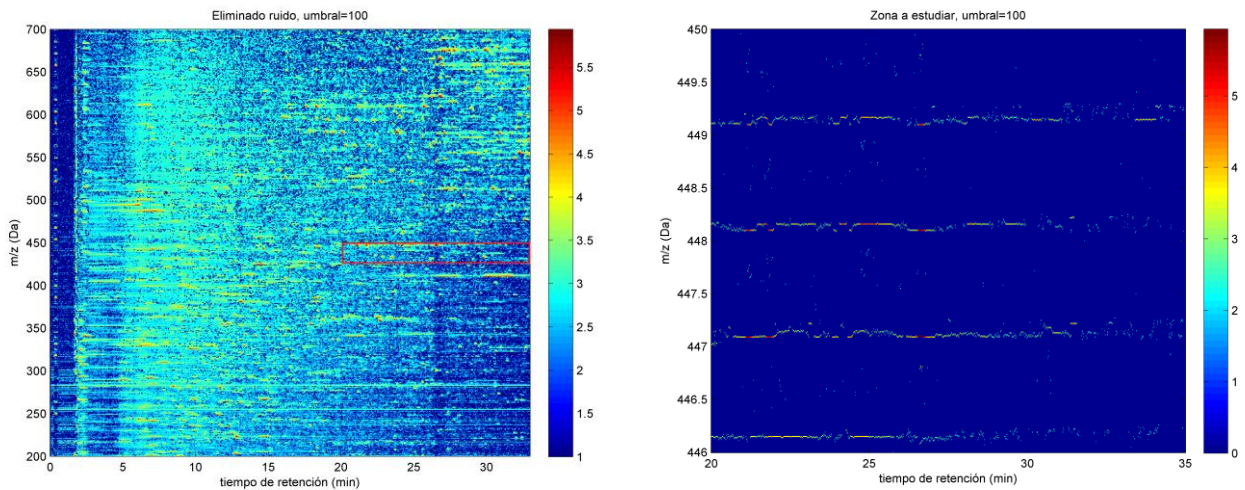


Figura 53. Ejemplo de aplicación del método de *preclustering*. A la izquierda el espectrograma de baja resolución (1 s, 1 Da) de un extracto de pimienta con reducción de ruido de umbral 100. El cuadrado rojo indica la zona a estudiar que comprende los intervalos en tiempo [20 min, 35 min] y masa [446 Da, 450 Da] y se representa a la derecha su espectrograma de alta resolución (1 s, 10 mDa).

Como primer paso, se aplica el "preclustering: fase 1" con $T_{\text{resol}} = 300$ s, $dt_R = 20$ s y $dm/z = 20$ mDa y se obtiene cajas de anchura T_{resol} , separadas por los gaps en m/z . En la Figura 54, se muestran los valores de cada punto (t_R , m/z) medidos en cajas. Se representan las cajas con el mismo color si se tocan y distinto color si están lo suficientemente separadas, según los umbrales establecidos. El conjunto de cajas del mismo color define un *cluster* y se le asigna un índice.

Se consigue así, una representación completa de los datos reagrupados en *clusters*, unidades pequeñas que se pueden tratar separadamente. La separación entre *clusters* se observa que es segura y todos los puntos pertenecen a algún *cluster* y sólo a uno. Se observa que hay muchos *clusters* muy pequeños (con pocos puntos) que probablemente son ruido y que en una fase posterior son fácilmente eliminables simplemente poniendo un umbral de

intensidad o área. En este proceso de *preclustering*, de momento no se elimina ninguno.

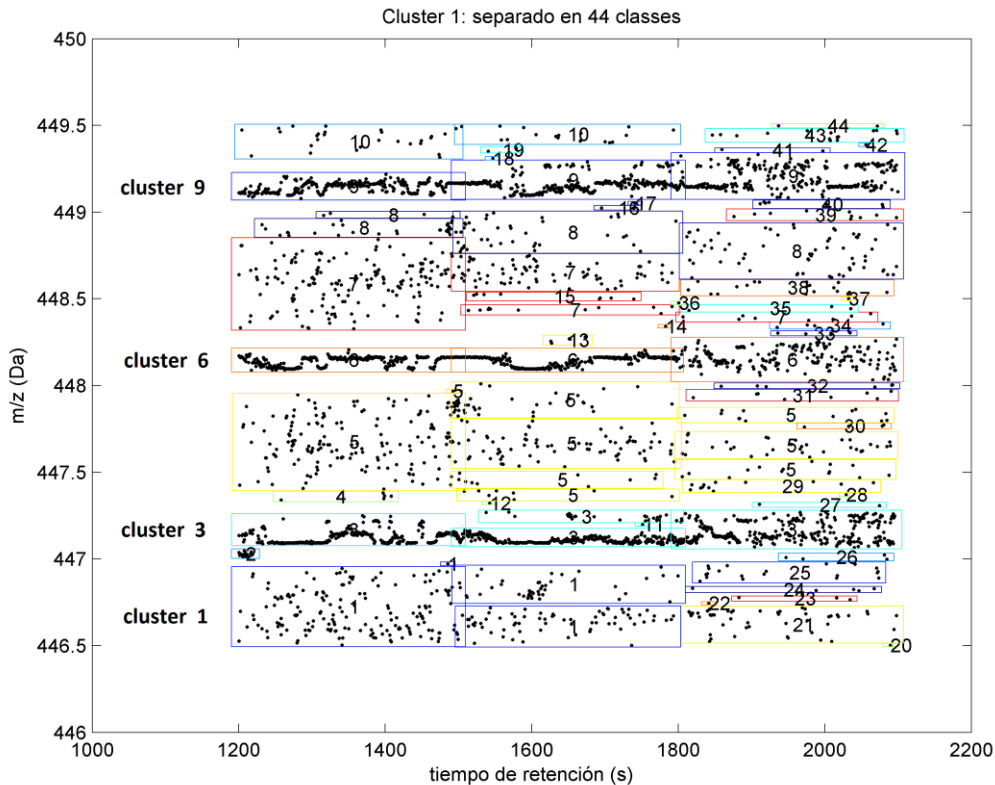


Figura 54. *Preclustering*: fase 1, basado en cajas de $T_{\text{resol}}= 300$ s, $dt_R= 20$ s y $dm/z= 20$ mDa. Aplicado a la zona de estudio del espectrograma del extracto de pimienta de la Figura 53 (izquierda).

También se observa que en este primer *preclustering* fase 1, se han formado *clusters* muy grandes con intervalos de tiempo de 600 y 800 s, por lo que evidentemente no se trata de una sola especie iónica, y tendrán que ser separados en los procesos posteriores. Este primer *preclustering* acota el problema y facilita el problema de la rapidez de cálculo.

Después del primer *preclustering* realizado, se vuelve a aplicar una o varias etapas de *preclustering* basadas en cajas pero con ventanas de tiempo menores. Con esto, se consigue que la mayoría de los *clusters* anteriores se puedan

dividir. En la Figura 55, se muestran los resultados de aplicar un "preclustering: fase 2" con $T_{\text{resol}}=20$ s y umbrales $dt_{\text{R}}=20$ s y $dm/z=20$ mDa sobre el *cluster* 1 de la Figura 54. Se observa que este *cluster* se ha dividido en muchas cajas más pequeñas y que las cajas del mismo color definen nuevos *clusters*. En total, el *cluster* 1 se ha dividido en 86 *clusters* nuevos.

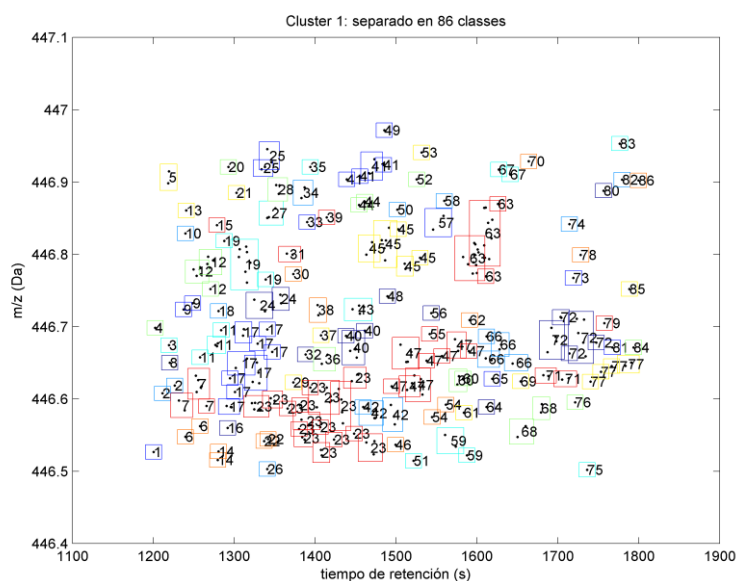


Figura 55. *Preclustering*: fase 2 basado en cajas de $T_{\text{resol}}=20$ s, $dt_{\text{R}}=20$ s, $dm/z=20$ mDa, sobre el *cluster* 1 de la Figura 54.

Paso 2 del *clustering* de líneas. *Clustering* de puntos.

Una vez aplicado el *preclustering* basado en cajas, se tiene los datos LC-MS estructurados en *clusters* (que son pocos), en lugar de puntos (t_{R} , m/z , I) (que son muchos). Para analizar cada *cluster* generado en el *preclustering* hay que abordar el problema desde otro enfoque. Como la complejidad de analizar cada *cluster* separadamente ahora es pequeña, se puede realizar un *clustering* basado en los puntos de cada *cluster*. Esto no hubiera sido posible debido al coste computacional si se aplicara directamente a los puntos del LC-MS sin el *preclustering*.



En el *clustering* de puntos (basado en los conceptos de lo que se conoce en la bibliografía como método intersection box)¹³⁹ para cada valor de t_R considera su espectro de línea (m/z , I) y asigna inicialmente un *cluster* a cada punto (t_R , m/z , I) del espectro de línea. Después, se hace un barrido en el eje t_R , de manera que para cada t_R , se determina la distancia que separa cada punto de su espectro de línea con los puntos del tiempo anterior t_R-1 . Si están a una distancia menor que el umbral de masa y tiempo establecido, se une al *cluster* anterior más cercano y si no lo cumple, se define un nuevo *cluster*. En este *clustering*, la cercanía a los puntos vecinos juega un papel importante. Los umbrales que se utilizan son más pequeños, $dt_R = 10$ s y $dm/z = 10$ mDa. En este caso, la relación de distancia puede extenderse más allá de la muestra adyacente y hay divisiones no cartesianas de los *clusters*, tal y como se aprecia en la Figura 56.

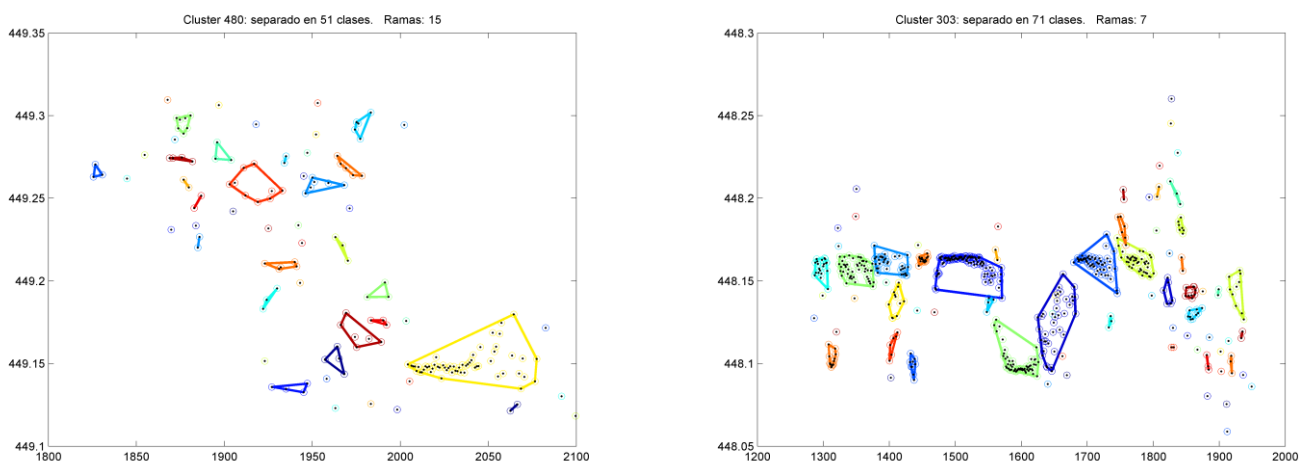


Figura 56. Resultado de aplicar el *clustering* de puntos sobre los *clusters* 6 y 9 de la Figura 55, que habían quedado sin dividir por el *preclustering* fase 1 y 2. Se ha aplicado umbrales de $dt_R = 10$ s y $dm/z = 10$ mDa.

¹³⁹ Kirchner, M. *et al.*, 2011. libfbi : A C++ implementation for fast box intersection and application to sparse mass spectrometry data. *Bioinformatics*, 27, pp.1166–1167.



Se aprecia que este tipo de clustering divide en varios clusters lo que no han podido dividir los clustering de la fase 1 y 2. En este punto ya es muy grande el número de clusters generados y en su mayoría contienen sólo una especie iónica o ruido (fácilmente eliminable), pero quedan algunos clusters con más de una especie iónica que es preciso separar en la siguiente fase del proceso.

Paso 3 del clustering de líneas. Detección de picos cromatográficos.

Una vez que el espectro está dividido en *clusters* muy pequeños, para algunos de ellos habrá terminado el proceso de división. Pero para los que no, se procede a realizar la detección de picos cromatográficos dentro de cada *cluster*. Esto se realiza detectando los máximos y mínimos locales del perfil $I(t_R)$ de cada *cluster*. Una vez detectados, se realiza la división en los mínimos locales y habrá tantos *cluster* o compuestos como máximos cromatográficos se localicen. Esta detección de picos cromatográficos es similar a la que se realiza en el método convencional, pero en este caso, se aplica al cromatograma que corresponde con un *cluster*^{121,122}.

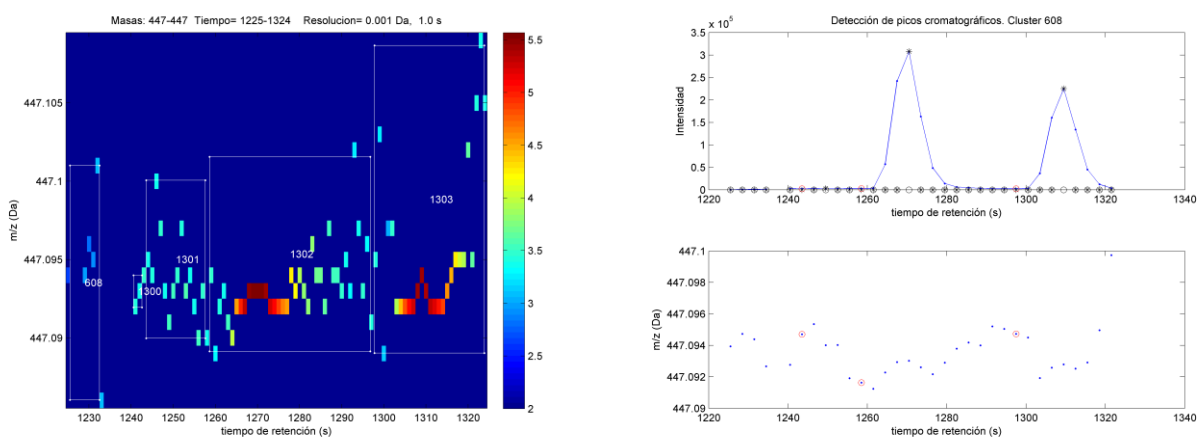
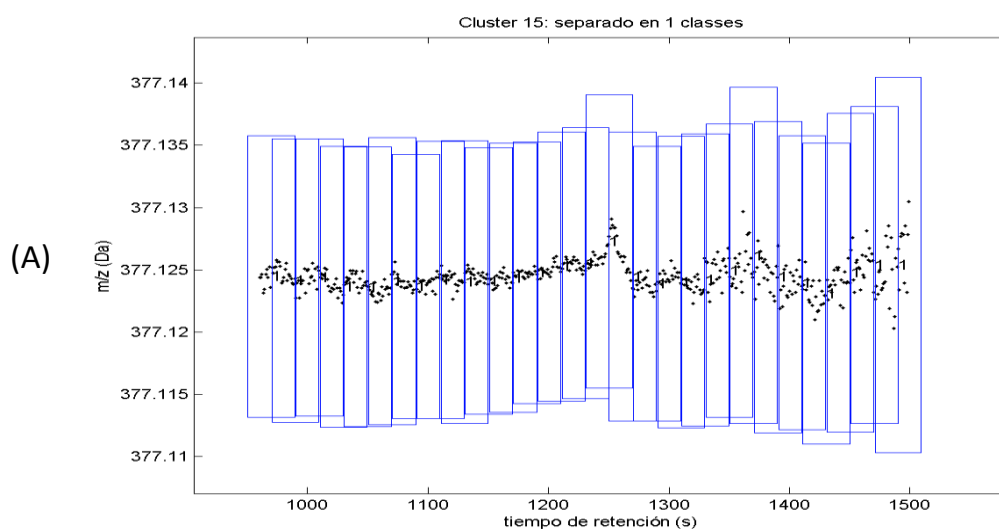


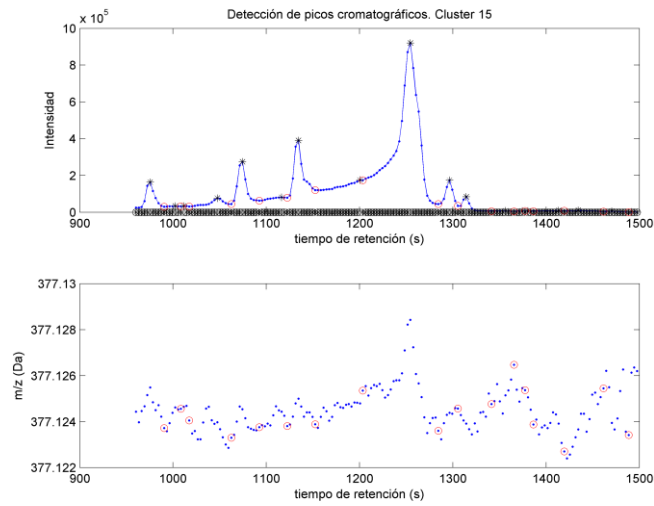
Figura 57. Clustering por detección de picos cromatográficos: resultado de aplicar la detección de picos cromatográficos a una parte del *cluster* 3 de la Figura 56, que no había podido dividirse por *preclustering* fases 1 y 2, o por *clustering* de puntos.



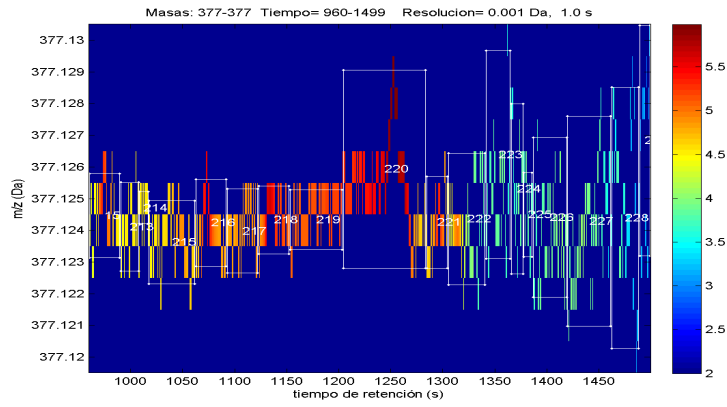
Como se ha mencionado, el método convencional tiene sus riesgos de dividir especies iónicas, pero si se realiza previamente el *preclustering*, ese riesgo es mucho menor. En la Figura 57, se muestra un ejemplo de un *cluster* que no se ha podido dividir mediante el *preclustering* fases 1 y 2, ni por el *clustering* de puntos y que con la detección de picos si ha sido posible. Se trata de una parte del *cluster* 3 de la Figura 56. En el cromatograma EIC de la Figura 57, se observan dos picos cromatográficos muy claros, así como los máximos y mínimos locales que se han detectado por separación del *cluster* inicial en 5 *clusters*. De estos 5 *clusters*, probablemente sólo 2 de ellos, los centrados en $t_R=1270$ s y $t_R=1310$ s y con $m/z=447.089$ Da y $m/z=447.094$ Da, respectivamente, correspondan a especies iónicas diferentes y los otros tres sean ruido.



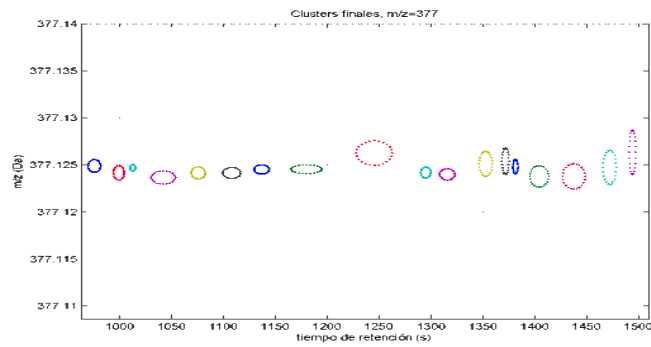
(B)



(C)



(C)



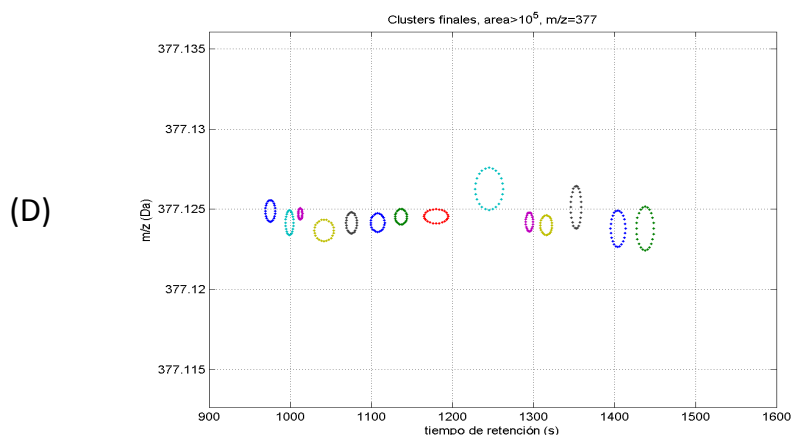


Figura 58. Resultado de aplicar el *preclustering* basado en cajas dos veces con $T_{\text{resol}}=300$ s y después $T_{\text{resol}}=20$ s, y el método de *clustering* de puntos. Se ha realizado un zoom centrado en $m/z=377$ Da del espectro de masas de un extracto fenólico de aceite de oliva (A). Método de *clustering* de detección de picos cromatográficos para el mismo caso que A (B).

Espectrograma para la zona estudiada y separación de los distintos *clusters* por el método de detección de picos cromatográficos (C). Representación esquemática de los *clusters* finales (D).

Representación esquemática de los *clusters* finales filtrados en área $>10^5$ (E).

En la Figura 58A, se muestra otro caso extremo en el que no se ha podido separar en *clusters* por ninguno de los métodos de *preclustering* ni *clustering* de puntos, y queda resuelto mediante la detección de picos cromatográficos. Corresponde a un *cluster* de un extracto de aceite de oliva. Como se aprecia en la figura de arriba, antes de la detección de picos cromatográficos, había un único *cluster*. Al aplicarle la detección de picos, en la representación de $I(t_R)$ (Figura 58B) se detectan máximos locales (marcados con círculos) y los mínimos locales (marcados con "x"). La Figura 58C es la misma representación que Figura 58A, donde se representa la relación m/z vs. t_R y los máximos y mínimos locales. En la Figura 58D se muestran los rectángulos con los nuevos *clusters* obtenidos al separar en los valores de los mínimos locales. Y por último, en la Figura 58E se ha realizado una representación esquemática en forma de elipses de colores de los *clusters* formados. Se ha representado cada *cluster* como una elipse con un

semieje dado por la desviación estándar de la masa y el otro semieje por la desviación estándar del tiempo de retención. Cada *cluster* representa una posible especie iónica distinta. En este caso, muchos de ellos tienen igual relación m/z (salvo desviaciones) pero distintos tiempos de retención, por lo que es posible que algunos de ellos sean isómeros de la misma especie iónica, y otros (con área muy inferior) puedan tratarse de ruido. Como se ha descrito anteriormente, este método de detección de picos cromatográficos tiene sus riesgos de generar falsos positivos, por lo que es necesario realizar un análisis posterior. Como primer paso para eliminar posibles *clusters* "falsos", se puede filtrar en área, quedando aquellos de mayor área, como se muestra en la Figura 58E.

III. Procesamiento de los datos obtenidos con el *clustering* de líneas.

- **Cálculo del área cromatográfica.**

Una vez que están definidos los puntos (t_R , m/z , I) asignados a cada *cluster*, se pueden determinar las variables que definirán cada *cluster*: tiempo de retención y su desviación estándar; m/z exacta y su desviación estándar, y una estimación del área del *cluster*.

La medida del área o de la altura del pico cromatográfico, es el factor más importante a la hora de realizar un análisis cuantitativo. Es muy cómodo utilizar la altura del pico para la cuantificación, si bien únicamente proporciona una exactitud aceptable en cromatogramas que presenten picos agudos, estrechos, claramente bien definidos y muy simétricos.

Existen multitud de métodos que permiten medir áreas bajo picos cromatográficos, uno de los más utilizados es el de "altura por la mitad de la anchura", el área se calcula como el producto de la altura por la anchura a la



mitad de la altura. Da buenos resultados para picos con forma aproximadamente gaussiana, pero los resultados son menos satisfactorios para picos no simétricos o para picos pequeños y anchos.

El método de cálculo de áreas más utilizado es mediante los integradores electrónicos: detecta el comienzo y el final de cada pico cromatográfico, integra digitalmente el área bajo la curva y corrige automáticamente la línea base. El problema de los integradores se plantea cuando hay picos parcialmente resueltos, ya que en los casos extremos llegan a integraciones disparatadas. Este problema se resuelve mediante la utilización integradores programados personalmente de manera que permitan al usuario tanto la visualización y supervisión de la forma en que se han integrado los picos, como la posibilidad de modificar la forma de la integración en caso de ser necesario, de forma manual. Esto requiere mucho tiempo y esfuerzo. El cálculo del área para los *clusters* obtenidos con el método desarrollado de *clustering* de líneas se realiza de forma automática y sin tener en cuenta el perfil del pico cromatográfico como en el método convencional. El área asociada al *cluster* k de una muestra n , se ha calculado como la suma de las intensidades de todos los puntos $(t, m/z, I)$ pertenecientes al *cluster* k , C_k :

$$A_n(k) = \sum_{j \in C_k} I_j \quad (\text{Ecuación 5})$$

- **Ficheros de resultados: listado de posibles compuestos.**

Después del tratamiento de las señales de LC-MS realizado: *preclustering* repetido, *clustering* de puntos y detección de picos cromatográficos, se obtiene un fichero de resultados con una lista de todos los *clusters* generados. Cada *cluster* está definido por un valor de tiempo de retención y masa exacta promedios, con sus desviaciones y el área o suma de intensidades de los puntos $(t, m/z, I)$ que pertenecen al *cluster*.



Los valores de tiempo de retención y masa exacta (m/z) y sus desviaciones estándar para cada *cluster* se calculan como medias pesadas en la intensidad sobre todos los puntos (t , m/z , I) pertenecientes al *cluster* C_k , con las siguientes fórmulas:

$$tr_n(k) = \frac{\sum_{j \in C_k} I_j tr_j}{\sum_{j \in C_k} I_j} \quad (\text{Ecuación 6})$$

$$m/z_n(k) = \frac{\sum_{j \in C_k} I_j m/z_j}{\sum_{j \in C_k} I_j} \quad (\text{Ecuación 7})$$

$$\sigma^2_{tr_n}(k) = \frac{\sum_{j \in C_k} I_j (tr_j - tr_n(k))^2}{\sum_{j \in C_k} I_j} \quad (\text{Ecuación 8})$$

$$\sigma^2_{m/z_n}(k) = \frac{\sum_{j \in C_k} I_j (m/z_j - m/z_n(k))^2}{\sum_{j \in C_k} I_j} \quad (\text{Ecuación 9})$$

Los ficheros que contienen estos datos de los *clusters* son fáciles de manejar y ordenar por áreas, t_{R} , o m/z . Suelen ser ficheros de unos pocos MB y pueden llegar a tener del orden de miles de *clusters* o posibles especies iónicas. Como este volumen de *clusters* es muy elevado para proceder a la identificación, lo que se hace es ordenarlos por área y seleccionar aquellos *clusters* con mayor área. Dependiendo del umbral de área seleccionado, se obtendrán listas de candidatos a compuestos más o menos extensas.

En la Tabla 5, se muestra un ejemplo de un fichero de resultados para la zona del espectrograma del extracto vegetal de pimiento estudiada, comprendida en el intervalo de tiempo [1200 s, 2100 s] y de masa [446 Da, 450 Da]. Se parte de datos filtrados en ruido y a los que se le ha aplicado un umbral de intensidad de



100 cuentas. Para cada *cluster* obtenido mediante el método de *clustering* de líneas se indica el índice, el área, la masa exacta y su desviación estándar, el rango de masa (diferencia entre el valor máximo y mínimo de masa de cada punto ($t_{r,m/z,l}$) que forman el *cluster*), el tiempo de retención y su desviación estándar y el rango de tiempo (diferencia entre el valor máximo y mínimo de tiempo de los puntos que componen el *cluster*)).

Tabla 5. Ejemplo de fichero de resultados estadísticos para la zona de la muestra de pimiento estudiada, comprendida en el intervalo de tiempo [1200 s, 2100 s] y de masa [446 Da, 450 Da].

Cluster	Área (*1e3 cuentas)	mass (Da)	SE_M (mDa)	rangoM (mDa)	time (s)	SE_T (s)	rangoT (s)
1313	8444.4400	447.0990	1.9654	16.3879	1598.0616	3.9641	45.160
1299	2569.9040	447.0955	0.6227	12.5427	1270.3813	3.9080	38.134
1397	2544.9280	448.0992	0.7905	22.2473	1597.7466	4.0527	54.190
1300	1864.1400	447.0954	1.1945	19.6838	1309.5231	3.1425	26.091
1394	848.4560	448.1662	1.4500	15.9912	1508.4753	11.5718	52.186
1007	754.7200	448.1659	1.4697	19.8669	1488.8306	4.2172	26.091
690	649.0480	448.0987	1.1242	15.6555	1270.1715	3.1310	22.078
1361	519.3760	449.1010	1.5488	24.0784	1597.8415	3.4907	34.121
989	339.3240	448.1045	4.0812	13.7329	1310.7792	2.3122	11.039
1385	336.7880	448.1610	5.3682	18.6157	1304.4734	1.6833	11.039
1005	269.9440	448.1654	1.0100	7.6904	1450.4768	2.5989	14.051
1358	154.4240	449.1695	1.8821	12.2375	1505.5549	7.5275	29.103
1357	145.5960	449.1690	3.1630	29.6326	1489.2530	3.5436	17.059
699	122.7360	449.1006	1.4799	9.9182	1269.9763	2.2172	9.033
1374	115.7160	449.1489	2.2283	20.4163	1817.2840	4.0722	25.090
1402	113.3840	448.1655	1.4164	6.6528	1695.9660	6.1983	26.093
631	99.1000	447.0964	2.6543	13.2751	1463.7478	2.4933	11.038
1312	85.8440	447.0963	2.3511	22.0337	1578.4058	4.1109	16.056
877	78.7600	447.2229	1.9157	15.4114	1885.9895	3.0520	25.089
1307	71.9920	447.0954	1.9395	20.0500	1434.8463	3.1734	17.060
.....							

En la Figura 59, se muestra el mismo ejemplo pero mediante una representación esquemática (mediante elipses) de todos los *clusters* que se han obtenido aplicando el *clustering* de líneas. Arriba se han representado todos los *clusters* (sin filtrar en área) y en las dos gráficas de abajo se representan los *clusters* filtrados por área $>10^4$ y área $>10^5$. Se puede apreciar que con estos umbrales, el número de *clusters* disminuye considerablemente, quedando sólo 78 *clusters* con área $>10^4$, 16 *clusters* con área $>10^5$ y sólo 4 *clusters* que cumplen la condición de tener área $>10^6$.



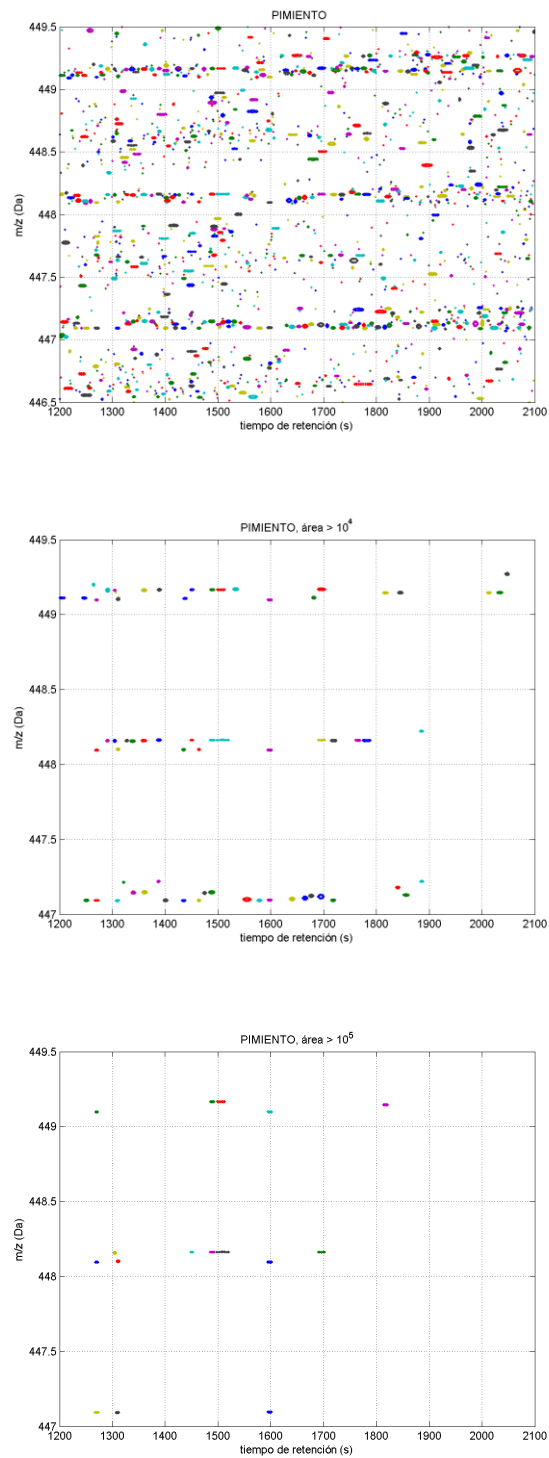


Figura 59. Arriba: representación de todos los *clusters* obtenidos, para la muestra estudiada de pimienta, sin eliminar ninguno. En medio: representación de todos los *clusters* obtenidos filtrados en área > 10⁴, y abajo: los *clusters* con área > 10⁵.

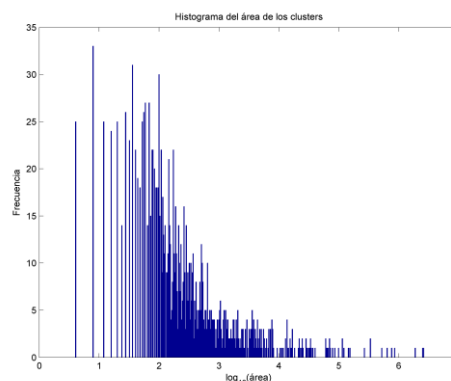


En la Figura 60, se muestra una tabla donde se indica el número de *clusters* generados en cada etapa del *clustering* de líneas, una tabla con el tiempo computacional que tarda cada paso y una figura con el histograma de las áreas de los *clusters*. Se observa que la mayoría de los *clusters* han finalizado su proceso de división después de aplicar el *preclustering* y el *clustering* de puntos y que un porcentaje pequeño de ellos ha necesitado resolverse finalmente mediante detección de picos cromatográficos.

También se muestra en la Figura 60 el histograma del área de todos los *clusters*, con lo que se puede observar la distribución de áreas. Un umbral de área de 100, muy conservador, permite tener la certeza de no estar eliminando ningún *cluster* asociado a una especie iónica y eliminaría mucho ruido. En este caso, se dispondría de 795 posibles especies iónicas a analizar. Para la caracterización, normalmente interesan los compuestos mayoritarios por lo que este número puede ser elevado y se podría elegir un umbral de área mayor. Pero para la identificación de compuestos bioactivos, sí supone una ventaja considerable disponer de todos las posibles especies iónicas pues, en principio, no se puede descartar ninguna por muy minoritaria que sea.



Clustering de líneas	Nº de clusters
Número inicial de puntos ($t_{R,m/z,l}$)	3782
<i>Pre-Clustering</i> fase1: $T_{\text{resol}}=300s$	113
<i>Pre-Clustering</i> fase2: $T_{\text{resol}}=20s$	374
<i>Clustering</i> de puntos	915
Detección de picos cromatográficos	1066
Umbral de área > 100	795
Umbral de área > 1000	286
Umbral de área > 10^4	78
Umbral de área > 10^5	16
Umbral de área > 10^6	4



Tiempo computacional

<i>Pre-Clustering</i> : fase 1 y 2	0.071 s	8%
<i>Clustering</i> de puntos	0.308 s	35%
Detección p. cromatogr.	0.164 s	18%
Estadística	0.087 s	10%
TOTAL	0.864 s	100%

Figura 60. A la izquierda: tabla con el número de *clusters* generados en cada una de las etapas del método de *clustering* de líneas. A la derecha: tabla con el tiempo computacional de cada etapa del *clustering* de líneas e histograma del área de los *clusters* generados. Corresponde a la zona de estudio comprendida entre los intervalos de tiempo [20 min, 35 min] y masa [446 Da, 450 Da] del extracto vegetal de pimiento, una vez que se ha quitado el ruido y se ha aplicado un umbral de intensidad de 100 cuentas.



Para este ejemplo, inicialmente se parte de una muestra con 3782 puntos $(t_R, m/z, I)$, que se convierten en una lista compacta y ordenada con datos estadísticos de unos pocos *clusters* (depende del umbral de área elegido) o posibles especies iónicas que forman la muestra. Además, este tipo de *clustering* propuesto invierte un tiempo computacional muy pequeño, como se muestra en la Figura 60. El tiempo computacional se ha medido con un ordenador que tiene las siguientes características: procesador Intel(R) Core(TM) 2 Duo CPU T6570, 2.10GHz, RAM 2,00GB y sistema operativo de 32 bits.

- **Identificación de compuestos a través de la huella espectrométrica.**

Una vez obtenida la lista de posibles compuestos, para facilitar la identificación de los mismos, se ha diseñado un algoritmo que agrupa a todos los *clusters* que tienen el mismo tiempo de retención (t_R) en familias. Este algoritmo se ha denominado **identificación de huellas espectrométricas o compuestos**. Cada familia de *clusters* representa la huella espectrométrica de un compuesto. Este tipo de agrupaciones en familias es muy útil ya que es muy usual encontrar muchas familias de especies iónicas en espectros de muestras vegetales. De esta forma, al detectar las huellas espectrométricas, se reduce el cálculo computacional y la comparación de los compuestos en diferentes muestras será más fácil. En las Figura 61 y en la Tabla 6, se muestran 4 huellas espectrométricas compuestas por la línea de un ión principal y sus isótopos para la zona estudiada de la muestra de pimiento. Estas 4 huellas espectrométricas se corresponden a 4 posibles compuestos de la muestra analizada para un umbral de área mayor de 10^5 .

Si en vez de analizar una zona reducida de una muestra, se analiza el espectro de la muestra completa, se obtendría la huella espectrométrica completa. Es decir, no sólo aparecería la línea del ión principal y sus variantes isotópicas sino también los posibles fragmentos, *clusters*, dímeros, etc. Para ilustrar las



características de una huella espectrométrica completa, se estudia un caso simple.

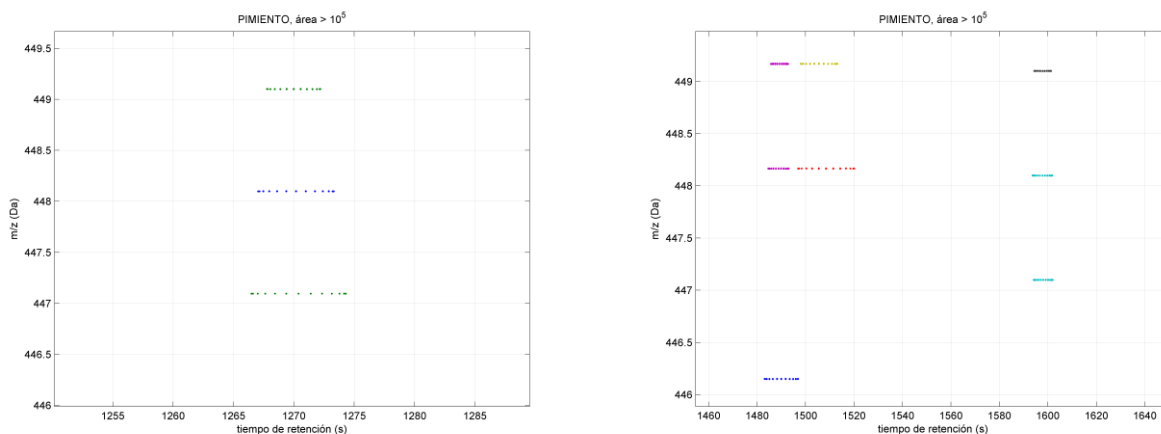


Figura 61. Representación de 4 huellas espectrométricas: ión principal y sus variantes isotópicas, una en cada figura.

Tabla 6. Compuestos calculados mediante el *clustering* de líneas y el procesado posterior para encontrar familias de especies iónicas: ión principal y sus isótopos, correspondientes a la zona de estudio de la Figura 59 donde se ha filtrado en área > 10⁵.

m/z (Da)	t _R (min)	Área(*10 ³)
COMPUESTO 1		
447.0955	21.2	2569.9
448.0987	21.2	649.0
449.1006	21.2	122.7
COMPUESTO 2		
446.1502	24.8	133.0
447.1502	24.8	29.4
COMPUESTO 3		
448.1662	25.1	848.5
449.1695	25.1	154.4
COMPUESTO 4		
447.0990	26.6	8444.4
448.0992	26.6	2544.9
449.1010	26.6	519.4



- **Ejemplo de aplicación del *clustering* de líneas a una muestra simple.**

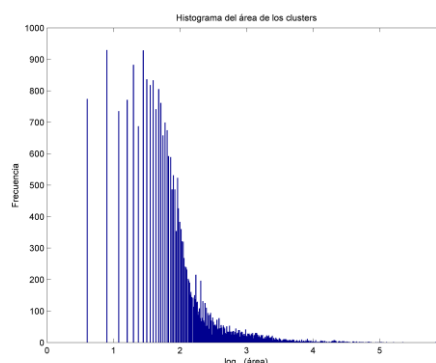
En este apartado, se estudia el caso de una muestra simple formada por la mezcla de 5 estándares conocidos, a la cual se aplica el *clustering* de líneas para obtener un listado de posibles compuestos. Con este ejemplo, se va a comprobar la robustez del algoritmo desarrollado y a exponer algunas consideraciones que hay que tener en cuenta a la hora abordar el estudio de las muestras complejas.

El *clustering* se realiza sobre los datos LC-MS filtrados en ruido y se le aplica un umbral de intensidad, $\text{umbral}_I=100$ cuentas. En la Figura 62 a la izquierda, se muestra una tabla con el número de puntos iniciales (t_R , m/z , I) (44.557 puntos sin ruido) y el número de *clusters* generados en cada etapa del *clustering* de líneas, y aplicando varios umbrales de área. A la derecha, se presenta una tabla con el tiempo computacional que tarda cada paso del *clustering* y el tiempo total, que para esta muestra es de aproximadamente 1 minuto.

También se muestra el histograma de las áreas de los *clusters* y se aprecia que un umbral de 100 es adecuado para eliminar los *clusters*, que posiblemente sean ruido de la muestra o generado por el mismo proceso de *clustering*.



Clustering de líneas	Nº de clusters
Número inicial de puntos ($t_R, m/z, I$) con ruido	165966
Número de puntos ($t_R, m/z, I$) sin ruido	44557
Pre-Clustering fase1: $T_{resol}=300s$	6873
Pre-Clustering fase2: $T_{resol}=20s$	12595
Clustering de puntos	18532
Detección de picos cromatográficos	20318
Umbral de área > 100	7849
Umbral de área > 1000	1347
Umbral de área > 10^4	215
Umbral de área > 10^5	14
Umbral de área > 10^6	1



Tiempo computacional

Pre-Clustering: fase 1 y 2	50.37s	65.5%
Clustering de puntos	2.79s	5.5%
Detección p. cromatogr.	4.42s	15.3%
Estadística	12.41s	13.7%
TOTAL	1.35 min	100%

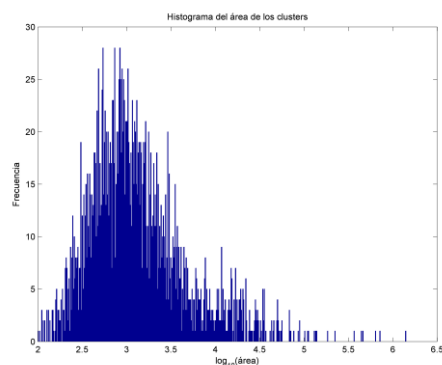
Figura 62. Ejemplo de una muestra simple mezcla de 5 estándares. A la izquierda: tabla con el número de *clusters* generados en cada una de las etapas del método de *clustering* de líneas. Se ha realizado con datos a los que se les ha eliminado el ruido y se ha puesto un umbral de intensidad de 100 cuentas . A la derecha: tabla con el tiempo computacional de cada etapa del *clustering* de líneas e histograma del área de los *clusters* generados.



Durante el proceso de *clustering*, en las sucesivas divisiones, se van obteniendo pequeños *clusters* (con pocos puntos), los cuales son ruido generado en el proceso de *clustering*. Estos *clusters* ralentizan el proceso de *clustering* recursivo, por lo que se ha desarrollado un algoritmo que los elimina con el siguiente criterio: los *clusters* con menos de tres puntos o con área menor de un umbral (umbral_área) se eliminan, ya que no corresponden a las características descritas de un compuesto. En la Figura 63 se presentan los resultados de eliminar *clusters* después de la ejecución de las distintas fases del *clustering*. Como era de esperar, se observa una disminución del tiempo computacional, siendo el resultado final de los *clusters* el mismo que en la figura anterior (Figura 62).



Clustering de líneas	Nº de clusters
Número inicial de puntos ($t_R, m/z, I$) con ruido	165966
Número de puntos ($t_R, m/z, I$) sin ruido	44557
<i>Pre-Clustering</i> fase1: $T_resol=300s$	6873
Elimina: puntos <3 y área <100	1107
<i>Pre-Clustering</i> fase2: $T_resol=20s$	5821
Elimina: puntos <3 y área <100	1191
<i>Clustering</i> de puntos	7117
Elimina: puntos <3 y área <100	1270
Detección de picos cromatográficos	3056
Elimina: puntos <3 y área <100	2580
Umbral de área > 1000	1347
Umbral de área $> 10^4$	215
Umbral de área $> 10^5$	14
Umbral de área $> 10^6$	1



Tiempo computacional

<i>Pre-Clustering</i> : fase 1 y 2	52.41s	86.1%
<i>Clustering</i> de puntos	2.77s	9.5%
Detección p. cromatogr.	6.06s	2.9%
Estadística	1.79s	1.5%
TOTAL	1.068min	100%

Figura 63. Mismo ejemplo que la Figura 62 pero eliminando ruido generado por el proceso de *clustering*.



En la Figura 64, se han representado los *clusters* obtenidos filtrados por $\text{área} > 10^4$ y, como se puede observar, se identifican las huellas espectrométricas de más de 5 compuestos. Se observan 8 huellas espectrométricas y además líneas continuas que se mantienen durante todo el tiempo de análisis de la muestra.

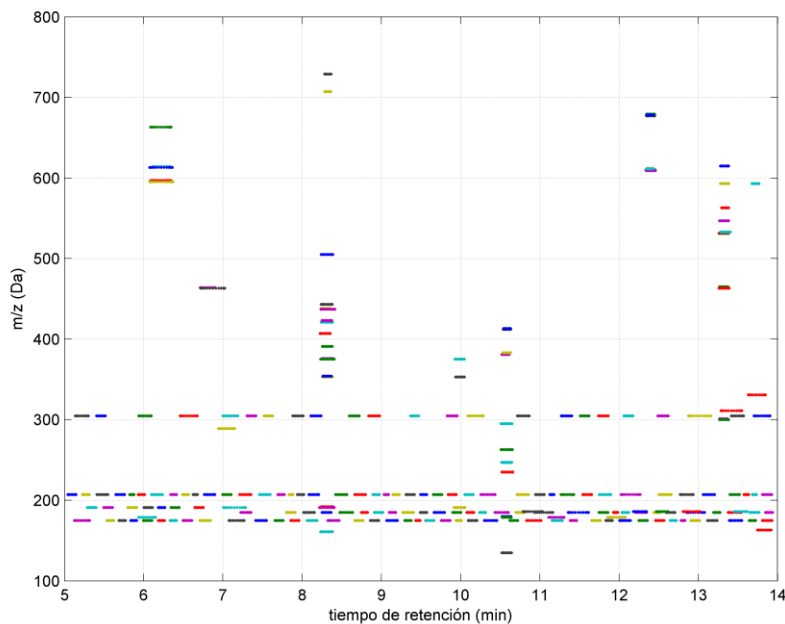


Figura 64. Representación de los *clusters* generados mediante el *clustering* de líneas, para una muestra compuesta de 5 estándares y filtrada para $\text{área} > 10^4$.

Después de aplicar el *clustering* de líneas, se aplica el algoritmo de identificación de huellas espectrométricas y se obtiene un listado con el que es más fácil identificar los posibles compuestos de la muestra. Este algoritmo también **elimina los *clusters* que aparecen en continuo**. El criterio para la eliminación es el siguiente: se consideran todos los *clusters* correspondientes a un intervalo de masa (definido por el parámetro $\text{dif_masa} = 10 \text{ mDa}$) y se suma la desviación estándar del tiempo de todos los *clusters* que se encuentran a lo largo de todo tiempo de análisis de la muestra. Si esta suma da un tiempo mayor o igual a un porcentaje del tiempo total ($\text{p_sigma_tr} = 25\%$), se eliminan. Es decir, si el análisis de la muestra dura 1800 s, se eliminan los *clusters* pertenecientes a un



intervalo de masa cuyas desviaciones estándar de tiempo de retención sumen más de 450 s.

En la Tabla 7 se muestra el listado de *clusters* que aparecen en continuo y que se han eliminado antes de realizar la selección de las líneas que forman la huella espectrométrica. También se muestra el listado de candidatos a compuestos ordenados por área y por tiempo de retención. Como se aprecia, se han identificado varios compuestos que eluyen en continuo y una vez analizadas las masas, es probable que estos compuestos sean componentes de la fase estacionaria, arrastrados por la fase móvil a través de la columna. Las masas encontradas en continuo son 185.0292 Da, 206.9742 Da, 174.9567 Da, 304.9152 Da, 288.9369 Da y 162.9832 Da.

En la Tabla 8, se muestra cada uno de los compuestos con las 10 líneas de mayor área de su huella espectrométrica. Para este umbral de área establecido (umbral_área_huella=15000 cuentas=15 Kcuentas) se han identificado 8 compuestos con su familia de líneas que eluyen en un mismo valor de t_R (que son los fragmentos, dímeros, *clusters*,...y cada uno con sus isótopos).



Tabla 7. Resultado de aplicar el algoritmo de identificación de compuestos a través de las familias de líneas o huellas espectrométricas a la lista de *clusters* obtenida con el *clustering* de líneas a una muestra simple con 5 estándares. Se ha utilizado un valor de umbral_área_huella= 15.000 cuentas. Se muestra los *clusters* eliminados, y los compuestos ordenados por área y por tiempo de retención. Para cada compuesto se especifica: la relación m/z, y su desviación estándar (en Da), el tiempo de retención y su desviación estándar (en minutos y segundos), el área (en Kcuentas= 1000 cuentas).

Fichero: [matriz15.bin](#)

Eliminados [45 clusters](#) alrededor de la masa [185.0292 Da](#): [t_tot=148.3 s](#) de [535.5 s](#)

Eliminados [43 clusters](#) alrededor de la masa [206.9742 Da](#): [t_tot=144.4 s](#) de [535.5 s](#)

Eliminados [45 clusters](#) alrededor de la masa [174.9567 Da](#): [t_tot=145.1 s](#) de [535.5 s](#)

Eliminados [39 clusters](#) alrededor de la masa [304.9152 Da](#): [t_tot=143.6 s](#) de [535.5 s](#)

Eliminados [42 clusters](#) alrededor de la masa [288.9369 Da](#): [t_tot=137.8 s](#) de [535.5 s](#)

Eliminados [46 clusters](#) alrededor de la masa [162.9832 Da](#): [t_tot=139.7 s](#) de [535.5 s](#)

Número de [clusters](#) iniciales: [2320](#)

Número de [compuestos](#) finales: [8](#)

Número de [clusters](#) iniciales: [2320](#)

Comp.1 m= [463.0899Da](#) se(m)= [1.0mDa](#) t=[13.33min](#) t=[799.5s](#) se(t)= [3.9s](#) A=[1415.6Kc](#) tiene [53líneas](#)

Comp.2 m= [609.1485Da](#) se(m)= [1.0mDa](#) t=[12.40min](#) t=[743.9s](#) se(t)= [3.6s](#) A= [719.1Kc](#) tiene [45líneas](#)

Comp.3 m= [191.0569Da](#) se(m)= [0.9mDa](#) t= [8.31min](#) t=[498.7s](#) se(t)= [5.7s](#) A= [638.9Kc](#) tiene [63líneas](#)

Comp.4 m= [595.1329Da](#) se(m)= [1.9mDa](#) t= [6.22min](#) t=[373.3s](#) se(t)= [8.7s](#) A= [461.9Kc](#) tiene [33líneas](#)

Comp.5 m= [179.0360Da](#) se(m)= [0.7mDa](#) t=[10.58min](#) t=[634.9s](#) se(t)= [3.5s](#) A= [446.3Kc](#) tiene [101líneas](#)

Comp.6 m= [191.0570Da](#) se(m)= [1.0mDa](#) t= [9.98min](#) t=[599.1s](#) se(t)= [4.2s](#) A= [90.1Kc](#) tiene [44líneas](#)

Comp.7 m= [463.0885Da](#) se(m)= [2.8mDa](#) t= [6.87min](#) t=[412.3s](#) se(t)= [9.2s](#) A= [76.7Kc](#) tiene [16líneas](#)

Comp.8 m= [331.0464Da](#) se(m)= [3.7mDa](#) t=[13.74min](#) t=[824.3s](#) se(t)= [6.6s](#) A= [15.4Kc](#) tiene [34líneas](#)

Lista de [compuestos](#) ordenados por tiempo de retención

Comp.1 m= [595.1329Da](#) se(m)= [1.9mDa](#) t= [6.22min](#) t=[373.3s](#) se(t)= [8.7s](#) A= [461.9Kc](#) tiene [33líneas](#)

Comp.2 m= [463.0885Da](#) se(m)= [2.8mDa](#) t= [6.87min](#) t=[412.3s](#) se(t)= [9.2s](#) A= [76.7Kc](#) tiene [16líneas](#)

Comp.3 m= [191.0569Da](#) se(m)= [0.9mDa](#) t= [8.31min](#) t=[498.7s](#) se(t)= [5.7s](#) A= [638.9Kc](#) tiene [63líneas](#)

Comp.4 m= [191.0570Da](#) se(m)= [1.0mDa](#) t= [9.98min](#) t=[599.1s](#) se(t)= [4.2s](#) A= [90.1Kc](#) tiene [44líneas](#)

Comp.5 m= [179.0360Da](#) se(m)= [0.7mDa](#) t=[10.58min](#) t=[634.9s](#) se(t)= [3.5s](#) A= [446.3Kc](#) tiene [101líneas](#)

Comp.6 m= [609.1485Da](#) se(m)= [1.0mDa](#) t=[12.40min](#) t=[743.9s](#) se(t)= [3.6s](#) A= [719.1Kc](#) tiene [45líneas](#)

Comp.7 m= [463.0899Da](#) se(m)= [1.0mDa](#) t=[13.33min](#) t=[799.5s](#) se(t)= [3.9s](#) A=[1415.6Kc](#) tiene [53líneas](#)

Comp.8 m= [331.0464Da](#) se(m)= [3.7mDa](#) t=[13.74min](#) t=[824.3s](#) se(t)= [6.6s](#) A= [15.4Kc](#) tiene [34líneas](#)



Tabla 8. Compuestos obtenidos y 10 de las líneas de mayor área de su huella espectrométrica. Se especifica la masa (Da), el tiempo de retención (min), el área (Kcuentas) y el porcentaje relativo de área de cada una de las líneas que componen la huella espectrométrica respecto del ión de mayor área. En gris, el ión principal de las huellas espectrométricas.

Comp.1 m=595.1329Da t=6.2min A=461.9Kc				Comp.5 m=179.0360Da t=10.6min A=446.3Kc			
595.1329	6.2	461.90	100.00	179.0360	10.6	446.32	100.00
596.1361	6.2	135.92	29.43	135.0478	10.6	111.02	24.88
613.1433	6.2	69.03	14.95	262.9872	10.6	51.22	11.48
597.1384	6.2	37.44	8.10	180.0399	10.6	41.41	9.28
663.1163	6.2	33.20	7.19	383.0348	10.6	18.24	4.09
614.1498	6.2	16.99	3.68	247.0128	10.6	16.44	3.68
679.0862	6.2	9.29	2.01	234.9735	10.6	16.23	3.64
664.1222	6.3	6.05	1.31	381.0537	10.6	15.20	3.40
615.1432	6.3	3.01	0.65	412.9940	10.6	13.05	2.92
158.9801	6.3	2.70	0.58	295.0028	10.6	12.21	2.74
Comp.2 m=463.0885Da t=6.9min A=76.7Kc				Comp.6 m=609.1485Da t=12.4min A=719.1Kc			
463.0885	6.9	76.72	100.00	609.1485	12.4	719.06	100.00
464.0908	6.8	11.38	14.84	610.1521	12.4	221.73	30.84
242.9434	6.9	2.98	3.89	611.1545	12.4	57.02	7.93
434.8750	6.8	2.80	3.64	677.1342	12.4	50.11	6.97
158.9792	6.9	2.30	3.00	678.1373	12.4	17.48	2.43
218.9806	6.9	2.04	2.66	679.1235	12.4	12.16	1.69
465.1018	6.9	2.00	2.60	612.1557	12.4	10.63	1.48
531.0644	6.9	1.36	1.77	739.1055	12.4	8.51	1.18
248.9040	6.9	0.82	1.07	693.1039	12.4	7.71	1.07
547.0401	7.0	0.81	1.05	178.9778	12.5	6.82	0.95
Comp.3 m=191.0569Da t=8.3min A=638.9Kc				Comp.7 m=463.0899Da t=13.3min A=1415.6Kc			
191.0569	8.3	638.91	100.00	463.0899	13.3	1415.60	100.00
375.0687	8.3	185.18	28.98	464.0926	13.3	370.71	26.19
353.0871	8.3	89.24	13.97	465.0948	13.3	70.88	5.01
437.0392	8.3	55.27	8.65	531.0739	13.3	50.22	3.55
192.0603	8.3	50.30	7.87	300.0269	13.3	35.66	2.52
376.0713	8.3	34.59	5.41	547.0426	13.3	28.90	2.04
443.0520	8.3	24.88	3.89	301.0351	13.3	16.97	1.20
161.0249	8.3	16.84	2.64	563.0152	13.3	15.46	1.09
354.0900	8.3	16.39	2.57	532.0736	13.3	15.44	1.09
505.0218	8.3	15.54	2.43	615.0290	13.3	11.83	0.84
Comp.4 m=191.0570Da t=10.0min A=90.1Kc				Comp.8 m=331.0464Da t=13.7min A=15.4Kc			
191.0570	10.0	90.11	100.00	331.0464	13.7	15.38	100.00
375.0704	10.0	30.26	33.58	593.1552	13.7	11.78	76.57
353.0874	10.0	11.34	12.59	311.1688	13.8	8.39	54.56
437.0397	10.0	8.50	9.43	349.0576	13.7	5.85	38.02
192.0585	10.0	8.22	9.12	594.1593	13.7	3.43	22.29
178.9764	9.9	6.16	6.84	130.9674	13.7	2.36	15.34
376.0716	10.0	5.19	5.76	191.9477	13.7	2.13	13.86
190.9284	10.0	3.28	3.64	232.0440	13.8	1.76	11.44
443.0532	10.0	2.75	3.05	204.9987	13.7	1.68	10.90
423.0493	10.0	2.55	2.83	242.9412	13.8	1.62	10.53

Los compuestos 1, 3, 5, 6 y 7 identificados mediante el *clustering* de líneas se corresponden con los 5 estándares empleados para preparar la muestra, como se aprecia en la Tabla 9. Se han obtenido valores de masa experimental que se desvían del valor teórico en menos de 1 o 2 mDa. El compuesto 3 se ha identificado gracias a la huella espectrométrica ya que el ión principal de masa



$m/z=353.0862$ Da no es el ión con mayor área. Para este compuesto, el ión que representa su huella espectrométrica es el de mayor área ($m/z=191.0569$ Da), encontrándose el ión principal en tercer lugar en cuanto al valor de área. Esto es muy habitual, por lo que es importante, a la hora de identificar compuestos, estudiar detenidamente su huella espectrométrica.

Además el método de *clustering*, ha proporcionado tres posibles compuestos nuevos: el 2, el 4 y el 8. Estudiando su huella espectrométrica, posiblemente tengan valores de masa y tiempo: $m/z=463.09$ Da, $t_R=6.88$ min; $m/z=375.10$ Da, $t_R=10$ min y $m/z=331.04$ Da y $t_R=13.8$ min. En el cromatograma y el espectrograma de la Figura 65, se aprecian los 5 compuestos de los estándares y también los tres compuestos nuevos.

Tabla 9. Comparación entre el resultado de *clustering* de líneas y los compuestos reales de la muestra. Se indican los nombres de los compuestos fenólicos, la fórmula molecular de su ión, su masa exacta calculada con la fórmula molecular, y los valores de m/z y t_R experimentales obtenidos mediante el *clustering* de líneas.

Compuesto fenólico	Fórmula ión	Masa exacta (Da)	m/z exp. (Da)	t_R exp. (min)
Delfinidina-3-O-sambubioside	$C_{26}H_{27}O_{16}^-$	595.1304	595.1325	6.2
Ácido clorogénico	$C_{16}H_{17}O_9^-$	353.0878	353.0862	8.3
Ácido caféico	$C_9H_7O_4^-$	179.0349	179.0348	10.6
Quercetina-3-O-rutinósido	$C_{27}H_{29}O_{16}^-$	609.1461	609.1482	12.4
Quercetina-3-O-glucósido	$C_{21}H_{19}O_{12}^-$	463.0881	463.0893	13.3



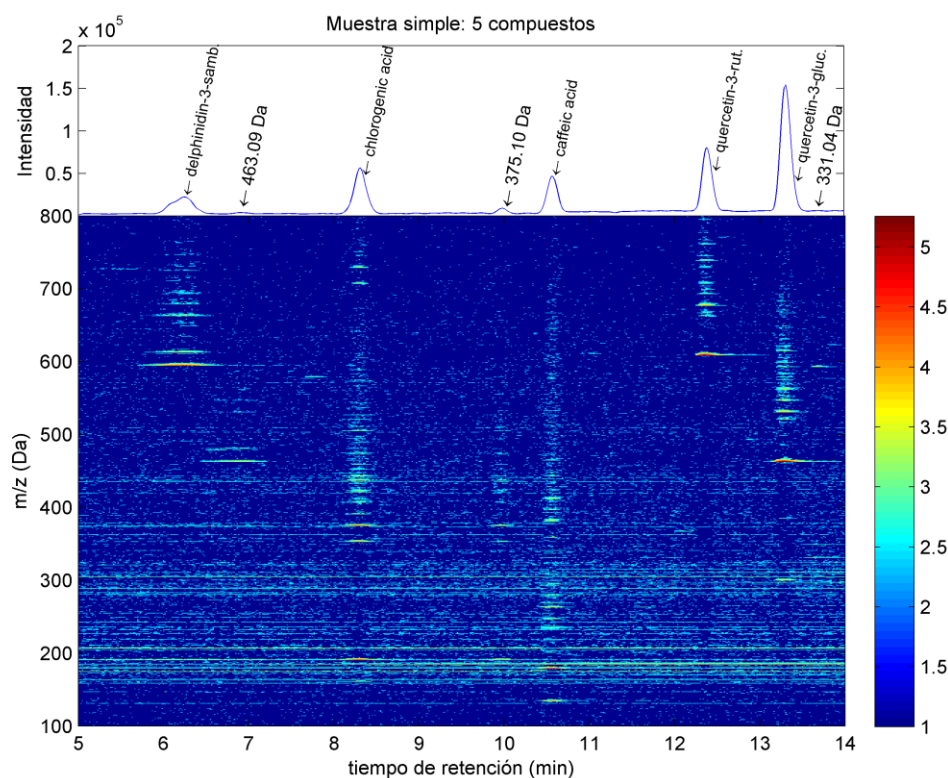


Figura 65. Cromatograma BPC y espectrograma de la muestra simple compuesta por 5 estándares. Aparecen 8 huellas espectrométricas que se corresponden con los 5 compuestos de la mezcla más tres compuestos nuevos: 463.09 Da y $t_R=6.9$ min; 191.0569 Da y $t_R=10.0$ min; 331.04 Da y $t_R=13.8$ min.

- **Procesamiento de una colección de muestras.**

La representación de los datos LC-MS basada en *clusters* es muy compacta: se sustituyen todos los puntos (t_R , m/z , I) de la muestra por un listado reducido de *clusters*. Cada *cluster* viene definido por los parámetros (t_R , σ_{t_R} , m/z , $\sigma_{m/z}$, área). Trabajar con ficheros que contienen listas de *clusters* es mucho más cómodo que con ficheros que contengan los puntos (t_R , m/z , I). Con estos ficheros de resultados que contienen los *clusters*, se puede realizar de forma mucho más fácil la identificación de compuestos, la comparación de diferentes muestras, así como realizar estudios estadísticos.

En los estudios donde se quiere comparar muchas muestras, normalmente del orden de 10 a 100 muestras, se obtendría un fichero por muestra con millones de puntos en cada uno de ellos. Con el método de *clustering*, se simplifica mucho la comparación de muestras ya que se hace de forma automática, comparando sólo unos miles o cientos de *clusters*. El número de *clusters* es controlable ya que se pueden ajustar los umbrales de intensidad y área para seleccionar más o menos *clusters*.

Además, en los estudios de colecciones de muestras es de especial importancia la rapidez con la que se puede realizar todo el proceso de *clustering*. El método de *clustering* de líneas, aplicando primero la reducción de ruido, los umbrales de área o intensidad adecuados y la eliminación de puntos espurios, puede reducir considerablemente el tiempo de tratamiento de los datos.

Para una colección de muestras, se genera una matriz donde cada columna corresponde a cada una de las N muestras, y cada fila corresponde al área del ión representante de cada uno de los compuestos identificados mediante el *clustering* de líneas y la identificación de huellas espectrométricas. Como se ha visto, el ión representante de un compuesto es el ión con mayor área de todos los que componen la huella espectrométrica y no siempre se corresponde con el ión principal. De esta forma, se dispone de una matriz de áreas A, donde el elemento $A_{i,j}$ representa el área del compuesto i para la muestra j . A partir de esta matriz, se realizan todos los cálculos estadísticos para cada caso particular.

- **Tabla con los valores de los parámetros del *clustering* de líneas.**

A continuación, se describen todos los parámetros que intervienen en el método de *clustering* de líneas y los valores que se han determinado como adecuados para las muestras vegetales analizadas en el desarrollo experimental de la tesis. Algunos de los valores de estos parámetros vienen determinados por



la observación de las características cromatográficas y espectrométricas de las muestras. Otros se han obtenido de un proceso de evaluación de los resultados sobre muestras simples. Estos parámetros son los que se han utilizado para realizar el *clustering* de líneas de las muestras de pimiento, *Cistus* y aceite de oliva.

Tabla 10. Parámetros aplicados en el *clustering* de líneas.

Descripción del parámetro	Valores por defecto
Umbral que se establece para limpiar los ficheros XML de ruido iónico	umbral_ruido_iónico=80
Umbral que se establece para filtrar los datos en intensidad (en cuentas), dejando pasar solo los valores mayores del umbral establecido	umbral_I=100
Tiempo de resolución de las cajas para el <i>preclustering</i> en la fase 1 (cajas grandes)	T_resol_fase1=300s
Tiempo de resolución de las cajas para el <i>preclustering</i> en la fase 2 (cajas pequeñas)	T_resol_fase2=20s
Umbral de tiempo de retención máximo de un <i>cluster</i> o especie iónica	umbral_tr=20s
Umbral de masa (m/z) máximo para la determinación de un <i>cluster</i> o especie iónica	umbral_m/z=0.010Da
Umbral de área para identificar huellas espectrométricas o compuestos (el valor de x depende del estudio concreto).	umbral_área_huella>10 ^x
Elimina <i>clusters</i> con umbral de área	umbral_área>100
Elimina <i>clusters</i> con menos de tres puntos	N_puntos=3
Eliminación de líneas en continuo	dif_masa=10mDa p_sigma_tr=25%

En el caso de querer aplicar el método desarrollado de *clustering* a otras muestras diferentes, sería necesario, ajustar los parámetros a las características cromatográficas y espectrométricas específicas de las muestras a analizar.



F) RELACIÓN ENTRE DATOS DE ESPECTROMETRÍA DE MASAS Y BIOACTIVIDAD.

Se ha desarrollado una herramienta para correlacionar datos de espectrometría de masas y datos de bioactividad, partiendo de colecciones de muestras vegetales complejas. El método de correlación propuesto sirve para la identificación de compuestos potencialmente bioactivos de dichas muestras. La identificación de candidatos a compuestos bioactivos a partir de una colección de muestras puede formularse de la siguiente manera. Se considera una colección de N muestras, (S_1, S_2, \dots, S_N) , consistente cada una de ellas en una mezcla compleja de compuestos que exhibe una actividad biológica determinada (y_1, y_2, \dots, y_N) . El objetivo del método propuesto es identificar qué compuestos son candidatos a ser responsables de la bioactividad observada. Para ello, se utiliza de cada muestra de la colección:

- ✓ El análisis químico (HPLC-MS). Estos análisis para las muestras S_n proporcionan la intensidad como una función del tiempo de retención (t_R) y la relación masa/carga (m/z). Debido a la técnica cromatográfica, cada compuesto eluye a un rango de tiempo específico. A este tiempo de elución, puede encontrarse en el espectro de masas una familia de picos de intensidad o huella espectrométrica, correspondiente a la masa del ión principal, sus variantes isotópicas, el patrón de fragmentación, etc.
- ✓ La medida de bioactividad. Para describir la bioactividad mediante un modelo básico, se ha de considerar que depende de la concentración de un único compuesto y que dicha dependencia entre la concentración (o área cromatográfica) y la actividad es lineal. El parámetro de bioactividad debe ser el adecuado de manera que se espere su variación lineal con la concentración del compuesto bioactivo. En algunos casos (como ocurre



por ejemplo, con el índice IC_{50}) no existe una relación lineal entre la concentración y el índice, y es necesario aplicar una transformación (mediante una función no lineal) para que se obtenga una dependencia lineal entre el nuevo parámetro de bioactividad y la concentración.

Teniendo en cuenta estos supuestos, se puede aplicar un análisis de correlación que involucre los datos HPLC-MS y los datos de la bioactividad correspondientes a las muestras.

En el estudio de correlación se encuentran involucradas N muestras y N valores de actividad de las muestras de la colección. Los datos de espectrometría de masas cubren rangos de cientos de Daltons con resoluciones del orden de 1 mDa, mientras que los tiempos de retención cubren rangos de varios minutos con resoluciones del orden de 1 s. Si los tiempos de retención o los valores experimentales m/z para un compuesto determinado están afectados por desviaciones durante el análisis HPLC-MS de las N muestras, las correlaciones para compuestos bioactivos pueden decrecer. Por esta razón, en lugar de utilizar los datos HPLC-MS crudos para el estudio de correlación, se ha de aplicar un tratamiento previo de los datos.

Mediante los métodos de *clustering* desarrollados, los datos HPLC-MS se transforman en un conjunto de *clusters* para cada muestra. Cada muestra queda representada mediante un conjunto de parámetros de área (se calcula el área de cada *cluster*). En el caso del *clustering* rectangular, cada parámetro de área se corresponde con un valor de $(t_R, m/z)$. En el caso del *clustering* de líneas, cada área representa el área de la línea del ión más representativo de cada compuesto identificado.

En cualquiera de los dos procedimientos de *clustering*, cada muestra viene representada por una lista de áreas. Dada una colección de N muestras, se



puede formar una matriz de datos de áreas donde cada fila se corresponde con los compuestos y cada columna, con las muestras. La identificación del compuesto bioactivo o candidato a compuesto bioactivo consiste en identificar qué compuesto está estadísticamente asociado a la actividad.

En el caso proceder de los datos tratados con el *clustering* rectangular, se realiza la correlación de las áreas de cada fila con la actividad. Las filas con mayor coeficiente de correlación indican los valores t_R y m/z correspondientes a los compuestos candidatos bioactivos. En el caso del *clustering* de líneas, las filas que proporcionen mayor correlación indican directamente los compuestos candidatos a bioactivos, contando con la huella espectrométrica de los mismos.

El modelo de correlación aplicado consiste en un modelo de regresión lineal simple entre el área $A_{i,n}$ de cada *cluster* i de cada muestra n de la colección ($n=1, 2, \dots, N$), y la bioactividad de la muestra y_n . Se realiza la correlación mediante la función afín:

$$y_n = a A_{i,n} + b \quad (\text{Ecuación 10})$$

donde b es la ordenada en el origen y a es la pendiente de la recta.

En el caso de que el número de muestras sea pequeño, se dispone de un número muy elevado de valores de áreas en comparación con el número de muestras. Por ello, en estos casos, se ha aplicado una correlación mediante la función lineal:

$$y_n = a A_{i,n} \quad (\text{Ecuación 11})$$

donde la ordenada en el origen es $b=0$, con lo que se deja un grado de libertad menor, imponiendo que la bioactividad sea nula cuando el área también lo sea.



Como resultado del análisis de correlación, se obtiene el coeficiente de correlación r (entre el área $A_{i,n}(t_{Ri}, m/z_i)$ y la bioactividad y_n) para cada *cluster* de las N muestras. Cuando r es cercano a 1, se puede decir que el compuesto correspondiente a dicho *cluster* está estadísticamente asociado con la bioactividad. Del mismo modo, la correlación más baja (r cercano a cero) correspondería con compuestos no asociados con la bioactividad. Por tanto, los *clusters* cuyo r alcanza valores cercanos a 1 permiten la identificación de los candidatos a compuestos bioactivos.

El análisis de correlación basado en *clustering* rectangular implica la realización de correlaciones entre los datos $A_{i,n}(t_{Ri}, m/z_i)$ y la bioactividad y_n de las N muestras, para cada valor $(t_{Ri}, m/z_i)$ del espectrograma. Típicamente con una resolución en masa de 1 Da y una resolución temporal de 1 s (para relación m/z y t_R respectivamente), un análisis con una duración de 30 minutos registrando m/z entre 100 Da y 1000 Da incluiría 1800 x 900 valores $(t_{Ri}, m/z_j)$, y por tanto, el estudio de correlación completo involucraría $1,62 \times 10^6$ análisis de correlación independientes. Hay que realizar 1.620.000 de correlaciones y, una vez identificadas las correlaciones altas, hay que identificar los compuestos, agrupando los *clusters* para los cuales la correlación es alta. Realizar millones de correlaciones es una tarea compleja y puede ocurrir que se obtengan correlaciones altas debido al azar (especialmente si el número N de muestras es bajo). Además, hay poca resolución en masa y se debe tener en cuenta el desalineamiento en tiempo y la calibración en masa.

También se han de tener en cuenta ciertos factores que pueden reducir el coeficiente de correlación de un $(t_R, m/z)$, siendo éste sin embargo, un compuesto bioactivo. Entre dichos factores, destaca el ruido de los datos HPLC-MS o de los datos de bioactividad, la dependencia no lineal entre la bioactividad y la concentración del compuesto bioactivo, la bioactividad influenciada por la



presencia de otros compuestos, etc. Asimismo, la correlación para picos asociados a compuestos no bioactivos debería ser baja, pero pueden encontrarse altas correlaciones (que conducen a falsos positivos) por diferentes razones. Si el número de muestras incluido en el estudio (N) es pequeño, y el número de picos cromatográficos (o compuestos en las muestras) es alto, la probabilidad de encontrar alta correlación debida al azar aumenta, de manera que el método podría proporcionar candidatos que en realidad no están asociados a la bioactividad. A veces, la concentración de varios compuestos puede asociarse al compuesto bioactivo (por ejemplo, debido a la propia fisiología vegetal de las muestras); en este caso, la correlación será también alta para el compuesto asociado, sin ser bioactivo. Por esta razón, el método propuesto no identifica el compuesto bioactivo, sino que proporciona una lista de candidatos a compuestos bioactivos.

En el caso de partir de los datos de *clustering* rectangular, debido a la ventana temporal, y dado que los compuestos eluyen durante un intervalo de tiempo, las regiones de alta correlación se ven como líneas en una representación de $r(t_{Ri}, m/z_j)$. Cuando un compuesto activo eluye a un intervalo de tiempo determinado, además de la línea m/z de alta correlación correspondiente a la masa de su ion, también se observan algunas líneas de alta correlación, correspondientes a los valores m/z asociados con este compuesto (su huella espectrométrica) debido a su fragmentación, patrón isotópico, etc. Así, en una representación del coeficiente de correlación en función del tiempo de retención y la m/z , $r(t_{Ri}, m/z_j)$ se pueden identificar fácilmente los candidatos a compuestos activos que aparecen en forma familias de líneas de alta correlación, todas ellas con diferentes valores m/z pero similares intervalos de tiempo de retención. De esta forma, la presencia de una línea aislada de alta correlación no estaría asociada con un compuesto activo. Tales correlaciones son espurias y vienen asociadas al ruido de los datos HPLC-MS. La probabilidad



de encontrar correlaciones de este tipo es mayor cuando el número de muestras N es bajo.

Tras la identificación de un *cluster* o familia de líneas de alta correlación, se puede identificar el candidato a compuesto bioactivo, comparando la información detallada proporcionada en forma de lista de candidatos con los datos de caracterización cualitativa de los extractos, previa o simultáneamente realizada.

La ventaja de utilizar el *clustering* de líneas radica principalmente en que el número de *clusters* es mucho más reducido, del orden de decenas o cientos y que es capaz de identificar el compuesto candidato a bioactivo directamente y de forma mucho más rápida.

Se ha aplicado la herramienta de correlación propuesta a los diferentes casos que conforman los tres bloques de resultados, correlacionando los datos analíticos de las diversas muestras (pimiento, *Cistus* y aceite) con datos de bioactividad (capacidad antioxidante, antimicrobiana y anticancerígena, respectivamente). Se ha realizado la correlación utilizando los datos procedentes de ambos métodos de *clustering*, resultando mucho más ventajoso el *clustering* de líneas. Por ello, es el *clustering* de líneas el que se presenta en los tres bloques de resultados de la tesis.

En el último bloque, en el que se aplica la herramienta de correlación a una colección de extractos polifenólicos de aceite de oliva con capacidad anticancerígena, se presenta el estudio de correlación para ambos métodos de *clustering*, el rectangular y el de líneas, con el objetivo de comparar ambas técnicas. Adicionalmente, en el bloque del aceite se propone la utilización de los compuestos candidatos para predecir la bioactividad. Cuando se identifica un candidato bioactivo, se puede calcular el área del pico cromatográfico a partir

del cromatograma de ion extraído. Al encontrar un coeficiente de alta correlación entre el área del pico y la bioactividad, se puede proponer entonces un modelo lineal para estimar la bioactividad a partir del área:

$$\hat{y}_n = a A_n + b,$$

donde \hat{y}_n es la bioactividad esperada para una muestra n , A_n es el área del pico cromatográfico seleccionado como candidato, b es la pendiente y a es la ordenada en el origen. De este modo, para una muestra m nueva, puede medirse mediante HPLC-MS el área para el pico cromatográfico A_m , y utilizando la Ecuación 10, se puede estimar o pronosticar el valor esperado para la bioactividad \hat{y}_m . La comparación entre los valores de la actividad pronosticada (\hat{y}_m) y los valores medidos experimentalmente (y_m) proporcionan un criterio para la validación de un candidato a compuesto bioactivo mediante la utilización de un grupo nuevo de muestras (distintas de los utilizadas para el estudio de correlación, la identificación de compuestos y la estimación de los parámetros a y b en la Ecuación 10).

A continuación, se presenta un experimento simple diseñado para evaluar el método de correlación propuesto en un ambiente donde las condiciones experimentales están bien controladas.

En el experimento se consideran 30 muestras preparadas en el laboratorio que contienen 5 estándares comerciales de compuestos fenólicos a diferentes concentraciones. La actividad estudiada para cada muestra es la absorbancia a una longitud de onda determinada, donde se conoce a priori que sólo la señal de absorbancia de uno de los compuestos va a ser significativa a esa longitud de onda. Se aplican en primer lugar los métodos de identificación de compuestos mediante *clustering* (rectangular y de líneas) a todas las muestras. Posteriormente, se aplica el método de correlación a 20 de las muestras,



elegidas de forma aleatoria (que constituirán el conjunto de muestras de entrenamiento). La correlación se hace entre el área de estos compuestos y la actividad, lo que permitirá identificar el/los compuesto/s candidatos a responsable/s de la actividad. Después, se aplica el modelo de predicción propuesto para evaluar la actividad de las 10 muestras restantes (que constituyen el conjunto de muestras de evaluación).

Los candidatos a compuestos activos determinados mediante la correlación utilizando el área de los compuestos calculados con *clustering* rectangular se van a comparar con los obtenidos utilizando el *clustering* de líneas. El objetivo es aplicar el método de correlación propuesto utilizando las dos herramientas de *clustering* a un caso sencillo, lo cual ayudará a entender mejor la aplicación del método en el caso de muestras reales complejas. Además, este estudio permitirá comparar los dos métodos de *clustering* desarrollados: rectangular y de líneas.

- **Diseño del experimento simple.**

El experimento incluye 30 muestras preparadas en el laboratorio que consisten en disoluciones en metanol de cinco estándares comerciales de polifenoles a diferentes concentraciones. Los compuestos utilizados c1, c2, c3, c4 y c5 se indican en la Tabla 11 y todos han sido adquiridos en Extrasynthèse (Genay, Francia). Las muestras se prepararon con concentraciones de los compuestos en un rango entre 0 y 8 µg/mL y un volumen total de metanol de 2 mL cada una. En la Tabla 12 se muestran las concentraciones utilizadas en cada una de las 30 muestras.



Tabla 11. Compuestos fenólicos de las 30 muestras: se indica el código identificador, el nombre, la fórmula molecular y la masa exacta.

Código	Compuesto fenólico	Fórmula molecular	Masa exacta (Da)
c1	delfinidina-3-O-sambubioside	C ₂₆ H ₂₈ O ₁₆	596.1384
c2	ácido clorogénico	C ₁₆ H ₁₈ O ₉	354.0957
c3	ácido caféico	C ₉ H ₈ O ₄	180.0429
c4	quercetin-3-O-rutinoside	C ₂₇ H ₃₀ O ₁₆	610.1540
c5	quercetin-3-O-glucoside	C ₂₁ H ₂₀ O ₁₂	464.0961

El parámetro de actividad considerado fue la absorbancia a 525 nm medida por espectrofotometría. La absorbancia para cada muestra se presenta en la Tabla 12 y viene definida como $\log_{10} (I_0/I_t)$ (como I_t disminuye exponencialmente con la concentración, la absorbancia aumenta linealmente con la concentración), donde I_0 y I_t son la intensidad incidente y transmitida, respectivamente. Entre los compuestos considerados para preparar las muestras, sólo c1 (que es una antocianina) presenta absorbancia significativa a 525 nm. De este modo, el método de correlación de la identificación de compuestos activos debe proporcionar un tiempo de retención y una m/z correspondiente al compuesto c1. Las últimas líneas en la Tabla 12 muestran el coeficiente de correlación r entre el parámetro de actividad y la concentración de cada compuesto involucrado en el experimento. Como se esperaba, el coeficiente de correlación es cercano a 1 sólo para el compuesto c1, y significativamente menor para el resto de compuestos (debido a la concentración usada para preparar las muestras, la correlación es también relativamente alta para c3).

Cada una de las muestras preparadas se analizó con HPLC-MS. Para la separación cromatográfica, se utilizó un RRLC Agilent 1200 con una columna Zorbax Eclipse (4,6 mm x 150 mm, con 1,8 μ m de tamaño de partícula). El flujo fue de 0,5 mL/min y el volumen de inyección de 10 μ L. Las fases móviles fueron:



(A) agua / acetonitrilo 90:10 (1 % de ácido fórmico) y (B) acetonitrilo. El gradiente lineal utilizado fue: 0 min , 5 % (B); 20 min , 40 % (B); 30 min, 5 % (B); 35 min, isocrático de 5 %(B). El equipo cromatográfico se acopló a un espectrómetro de masas TOF equipado con interfase ESI.

Los parámetros del espectrómetro de masas, operando en modo de ionización negativo, fueron: rango de masas, 50-1000 m/z; capilar, 4500 V; temperatura del gas, 190°C; flujo del gas de secado, 7 L /min; nebulizar presión de gas; presión del gas nebulizador, 1,5 bar; tasa de espectros, 1 Hz. Para el acoplamiento se incluyó un divisor de flujo 1:2. La calibración del TOF-MS se realizó mediante la inyección de *clusters* de formiato de sodio al principio de cada análisis.

Se asignaron al azar dos grupos de las muestras preparadas en este experimento. Se utilizó un grupo de 20 muestras (identificadas como "entrenamiento" en la Tabla 12) para realizar el análisis de correlación e identificar los compuestos candidatos a ser responsables de la actividad. El segundo grupo, formado por las 10 muestras restantes (identificadas como "evaluación"), se utilizó en el modelo de predicción para comparar la actividad medida con la actividad predicha usando muestras nuevas, no utilizadas previamente.



Tabla 12. Concentración de los 5 compuestos fenólicos (c1, c2, c3, c4 y c5) en cada una de las 30 muestras. Las dos últimas columnas representan los valores de absorbancia a 525nm de cada muestra y si la muestra se ha usado como entrenamiento o evaluación. Las últimas líneas muestran el coeficiente de correlación r entre la absorbancia y la concentración de los diferentes compuestos. Se ha calculado la correlación con las muestras usadas para entrenamiento ($r(\text{entrenam.})$), para evaluación ($r(\text{evalua.})$) y para todas las muestras ($r(\text{todas})$).

Muestra	Concentración ($\mu\text{g/mL}$)					Abs. 525 nm	Grupo
	c1	c2	c3	c4	c5		
1	0	0	0	0	0	0.001	Entrenamiento
2	0	0	0.1	0	0	0.007	Evaluación
3	2.0	0	0.1	0	0	0.050	Entrenamiento
4	0	2.0	0	2.0	2.0	0.010	Entrenamiento
5	2.4	1.6	0.4	2.0	0.8	0.043	Evaluación
6	0.4	1.6	0.8	1.2	0.4	0.066	Entrenamiento
7	0	0	0	0.1	0.1	0.004	Entrenamiento
8	1.6	1.0	0.6	1.2	1.4	0.028	Evaluación
9	2.4	1.0	1.4	0.8	0.6	0.043	Entrenamiento
10	4.0	0	2.0	1.0	0.10	0.059	Evaluación
11	2.0	2.0	1.0	0	2.0	0.037	Evaluación
12	2.0	2.0	0.8	1.6	1.0	0.026	Entrenamiento
13	0	2.0	0	2.0	2.0	0.003	Entrenamiento
14	4.0	2.0	2.0	2.0	0	0.074	Entrenamiento
15	4.0	2.0	2.0	2.0	2.0	0.071	Entrenamiento
16	400	200	300	0	200	0.066	Entrenamiento
17	6.0	2.8	2.0	0	3.2	0.096	Evaluación
18	6.0	2.0	4.0	1.0	0	0.096	Evaluación
19	4.0	3.0	0	4.0	4.0	0.065	Evaluación
20	0	3.0	0	4.0	2.0	0.006	Entrenamiento
21	0	0	0	0	4.0	0.006	Entrenamiento
22	0	4.0	0	4.0	0	0.005	Entrenamiento
23	6.0	3.0	3.0	3.0	3.0	0.113	Evaluación
24	6.0	0	3.0	0	0	0.101	Entrenamiento
25	4.8	2.0	3.6	3.0	3.0	0.077	Entrenamiento
26	4.8	2.8	2.0	3.2	3.6	0.082	Entrenamiento
27	4.0	0	2.0	4.0	4.0	0.066	Evaluación
28	8.0	2.0	4.0	0	0	0.140	Entrenamiento
29	8.0	3.0	4.0	3.0	2.0	0.135	Entrenamiento
30	8.0	4.0	4.0	2.0	2.0	0.142	Entrenamiento
$r(\text{entren.})$	0.9925	0.3682	0.9290	0.0339	0.1834		
$r(\text{eval.})$	0.9972	0.3983	0.6407	0.1186	0.3671		
$r(\text{todas})$	0.9921	0.3869	0.8935	0.0512	0.1076		

Abs.: absorbancia
entren.: entrenamiento
eval.: evaluación



- **Herramienta de correlación utilizando el *clustering* rectangular.**

Se procede a la identificación de los compuestos candidatos a ser responsables de la absorbancia a 525 nm, correlacionando las medidas de absorbancia y las áreas de los *clusters* generados mediante el *clustering* rectangular de los datos HPLC-MS.

Se aplicó el *clustering* rectangular a cada una de las 20 muestras de entrenamiento. Un ejemplo, correspondiente a la muestra 15 (muestra que contiene una concentración relativamente alta de los 5 compuestos) se explicó detalladamente en el apartado E2-II de la metodología. En la Figura 51 se representa el área $A_n(t_i, m_j)$ para cada par (t_i, m_j) con valores de resolución de 1s y 1 Da, calculada de acuerdo con la Ecuación 4. Se aprecian las familias de líneas espectrales correspondientes a los 5 compuestos que componen la muestra 15, además de tres compuestos identificados mediante el *clustering* que pueden ser isómeros, descomposición de alguno de los 5 compuestos mezclados inicialmente o contaminantes.

Se calculó para cada par (t_i, m_j) el coeficiente de correlación entre el área $A_n(t_i, m_j)$ y la actividad y_n de las 20 muestras de entrenamiento. El coeficiente de correlación $r(t_i, m_j)$ obtenido se ha representado en la Figura 66 (gráfica de en medio) mediante una escala de color (para aquellos valores de $r(t_i, m_j)$ mayores que 0.7). En la gráfica de arriba de la Figura 66 se ha representado el cromatograma BPC de la muestra 15. Se observan tres familias de líneas con valores altos del coeficiente de correlación alrededor de los tiempos de retención: 6.3 min, 6.9 min y 10.6 min. Las líneas de correlación tienden a presentarse en familias (correspondientes a las líneas de la huella espectrométrica de los compuestos candidatos a la actividad). Los picos aislados de correlación alta no corresponden a candidatos, son valores espurios de correlación alta debido al ruido. La familia de líneas alrededor de 6.3 min (la



mayoría de ellas con $r > 0.95$) están asociadas con el compuesto activo c1, la delfinidina-3-O-sambubioside. Las correlaciones observadas en 6.9 min corresponden a un compuesto (probablemente a un producto de descomposición del compuesto activo c1) con un área muy pequeña como se observa en el cromatograma BPC de la Figura 66. La familia de líneas alrededor de 10.6 min corresponde al compuesto c3 (ácido caféico) donde las correlaciones son inferiores a 0.9. Esta correlación no es consecuencia de la actividad sino que probablemente se deba a una cierta correlación entre la concentración de este compuesto y la concentración del compuesto activo c1 cuando se preparó la colección de muestras.

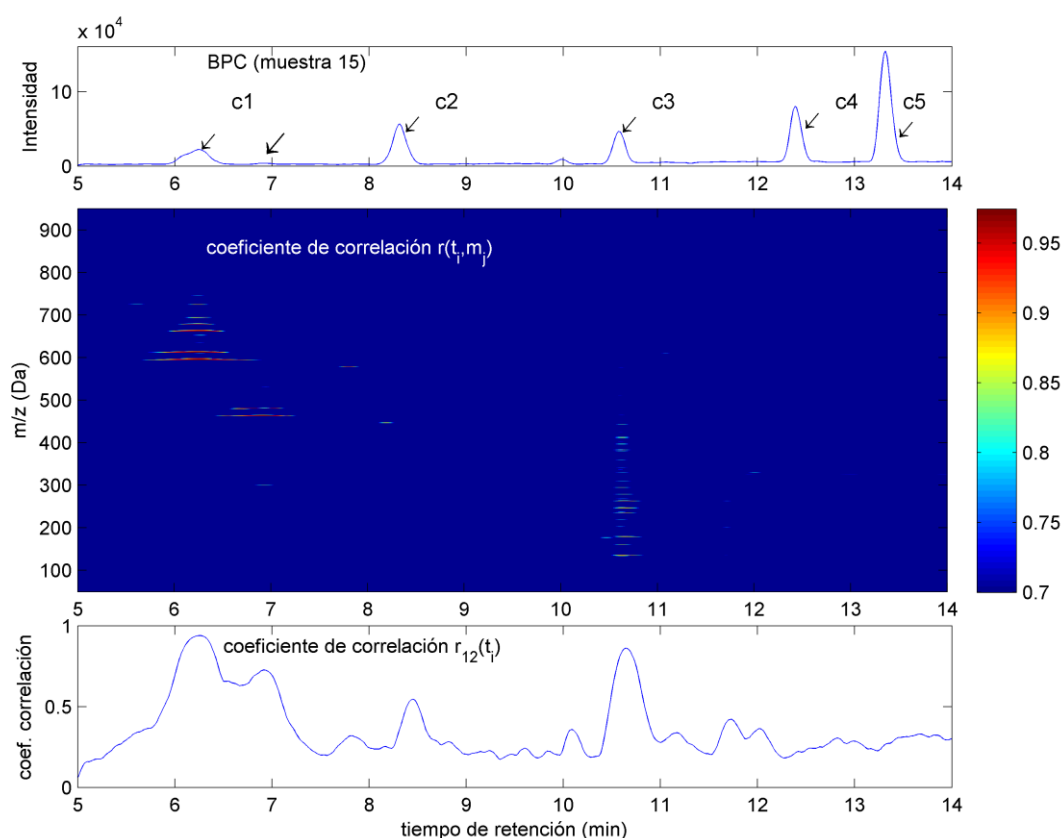


Figura 66. Arriba: cromatograma BPC para la muestra 15. Centro: coeficiente de correlación correspondiente a las 20 muestras seleccionadas para entrenamiento $r(t_i, m_j)$ entre el área $A_n(t_i, m_j)$ y la actividad y_n en función del tiempo de retención y la m/z. Abajo: representación de la función $r_{12}(t_i)$: media de los 12 valores mayores de $r(t_i, m_j)$ en el tiempo de retención t_i .

Para identificar de forma automática los candidatos a compuestos activos, se ha evaluado para cada tiempo de retención t_i el coeficiente de correlación promedio para los K valores más altos del coeficiente de correlación r , con la función $r_K(t_i)$:

$$r_K(t_i) = \frac{1}{K} \sum_{k=1}^K \rho(t_i, k) \quad (\text{Ecuación 12})$$

donde $\rho(t_i)$ son los coeficientes de correlación $r(t_i, m_j)$ en el tiempo de retención t_i ordenados en orden descendente, y K es el mínimo número de líneas que se espera encontrar en cada familia de líneas. Los tiempos de retención para los cuales el coeficiente de correlación promedio $r_K(t_i)$ presenta los valores más altos corresponden a los tiempos de retención de los candidatos más probables a ser compuestos activos. En la Figura 66 (abajo) se ha representado la función $r_K(t_i)$ para $K=12$. Los tiempos de retención de los mejores candidatos a compuestos activos son fáciles de identificar: $t_i=6.25$ min con $r_K(t_i)=0.939$; $t_i=10.6$ min con $r_K(t_i)=0.861$; y $t_i=6.9$ min con $r_K(t_i)=0.727$. Este resultado es consistente con las familias líneas de alta correlación observadas en la representación de $r(t_i, m_j)$. La identificación de los candidatos se basa en un análisis de las funciones $r(t_i, m_j)$ y $r_K(t_i)$ para cada tiempo de retención t_i . Se puede proponer como candidato para t_i el compuesto que cumpla: que $r_K(t_i)$ sea mayor que un umbral (se elige 0.7) y que el pico más abundante utilizado en la estimación de $r_K(t_i)$ tenga un coeficiente de correlación también mayor que el umbral establecido. Estos picos más abundantes proporcionan la m/z , el área y el coeficiente de correlación de un candidato que eluye en t_i cuando se cumplen las condiciones establecidas. La Figura 67 muestra la m/z , el área cromatográfica y el coeficiente de correlación de los picos más abundantes en función del tiempo de retención. Los tres candidatos obtenidos se indican en la Tabla 13.



Tabla 13. Compuestos candidatos obtenidos mediante la correlación entre el área de los *clusters* del *clustering* rectangular y la actividad de las 20 muestras de entrenamiento del experimento.

Compuesto	m/z (Da)	t_R (min)	$r(t_i, m_j)$
1	595	6.25	0.9746
2	463	6.9	0.9738
3	179	10.6	0.8732

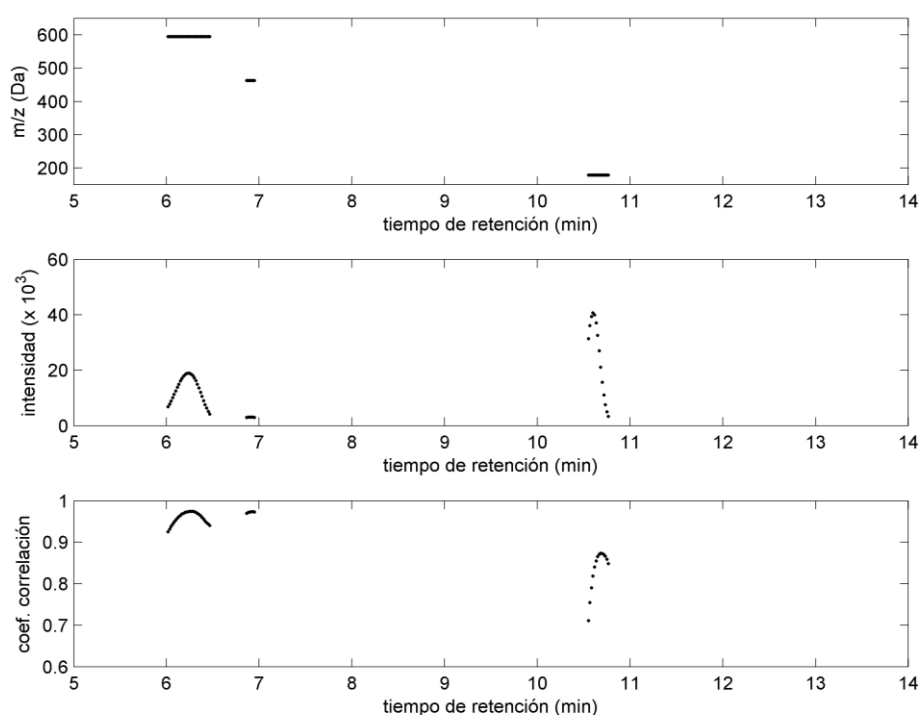


Figura 67. Identificación de los candidatos a compuestos activos. Se genera un punto cuando $r_{12}(t_i) > 0.7$ y el más abundante de los 12 picos seleccionados también muestra un coeficiente de correlación mayor de 0.7. En las gráficas se representa m/z (arriba), el área (centro) y el coeficiente de correlación (abajo) de los picos espectrales más abundantes en función del tiempo de retención cuando se cumplen las condiciones impuestas.

Adicionalmente, los K picos espectrales utilizados para la estimación de $r_K(t_i)$ proporciona los valores de m/z asociados (su huella espectrométrica) con el compuesto seleccionado como candidato en el tiempo t_i . Por ejemplo, para el



candidato que eluye a $t_r=6.25$ min, los valores m/z de 12 picos espectrales que forman su huella espectrométrica con mayor valor de r son: 595 Da, 596 Da, 613 Da, 597 Da, etc. En la Tabla 14 se indica su masa experimental, su área, su área relativa al máximo y el coeficiente de correlación asociado con los 12 picos para $t_r= 6.25$ min. Como se esperaba, el primer candidato obtenido es el compuesto de m/z 595.1306 Da y tiempo de retención 6.25 min, correspondiente a la delfinidina-3-O-sambubioside. La masa exacta de este compuesto es 595.13045836 Da y la desviación entre el valor exacto y el valor de masa del ion experimental es menor de 1ppm. También los valores de masa de los iones que componen su huella espectrométrica (*clusters*, fragmentos, variantes isotópicas,...) son también coherentes con las de la delfinidina-3-O-sambubioside. Se puede realizar un estudio similar con los otros candidatos a compuestos activos.

Tabla 14. Lista de candidatos a compuestos activos que eluyen en $t_r= 6.25$ min. Se indican las masas m/z experimentales y las áreas de los 12 picos espectrales con mayor coeficiente de correlación. Los picos están ordenados por intensidad.

Pico espectral	m/z experimental (Da)	m/z (Da)	Área muestra 15	Área relativa	r
1	595	595.1306	405084	100.00 %	0.974
2	596	596.1329	116872	28.85 %	0.970
3	613	613.1419	62584	15.45 %	0.971
4	597	597.1356	35816	8.84 %	0.969
5	663	663.1140	31868	7.87 %	0.959
6	614	614.1437	19732	4.87 %	0.954
7	664	664.1201	12820	3.16 %	0.948
8	679	679.0793	11364	2.81 %	0.931
9	598	598.1422	8284	1.05 %	0.928
10	665	665.1054	7956	1.96 %	0.890
11	615	615.1409	6608	1.63 %	0.892
12	695	695.0576	5072	1.25 %	0.879



- **Herramienta de correlación utilizando el *clustering* de líneas.**

En este caso, se correlacionan las medidas de absorbancia con las áreas del listado de posibles compuestos que proporciona de forma directa el *clustering* de líneas. Se ha aplicado el *clustering* de líneas a las 20 muestras de entrenamiento de la colección. En el apartado E2-III se estudia con detalle un ejemplo de aplicación del *clustering* de líneas para una muestra simple, que corresponde a la muestra 15 de la colección de muestras del experimento (Figura 64). Con este *clustering*, se obtiene directamente una lista de familias de compuestos que corresponden a las huellas espectrométricas de los 8 compuestos presentes en la muestra (Tabla 9).

En la Tabla 15 se muestran los compuestos candidatos a compuestos activos con sus correspondientes masas, tiempos de retención, área promedio y valores del coeficiente de correlación, resultado de aplicar la herramienta de correlación propuesta a las 20 muestras de entrenamiento. La Figura 68 es una representación de los coeficientes de correlación (con valores de $r > 0.5$) obtenidos para el representante de la familia de líneas (línea con mayor intensidad). Se representa el coeficiente de correlación mediante una escala de color (correlogramas). Como puede apreciarse, los tres primeros compuestos candidatos con valores de $r > 0.9$ coinciden con los obtenidos al aplicar el *clustering* rectangular y el resto de compuestos tienen un coeficiente de correlación muy bajo. Se obtiene que el compuesto responsable de la actividad es el compuesto de masa $m/z=595.1304$ Da con $t_R=6.23$ min, correspondiente al compuesto c1 (junto con el compuesto de masa $m/z=463.0863$ Da, $t_R=6.87$ min que está correlacionado con c1).

Dado que se obtienen resultados similares con ambos métodos de *clustering*, es preferible aplicar el método de *clustering* de líneas porque el cálculo es más directo y simple.



Tabla 15. Candidatos a ser responsables de la actividad, obtenidos mediante la correlación entre el área de los compuestos generados directamente mediante el *clustering* de líneas y la actividad de las 20 muestras de entrenamiento del experimento.

Compuesto	m/z (Da)	t _R (min)	área	r
1	595.1304	6.23	8345.2	0.9866
2	463.0863	6.87	1412.4	0.9712
3	179.0343	10.61	8847.4	0.9062
4	191.0561	8.33	11323.0	0.4187
5	191.0560	9.99	1638.5	0.2724

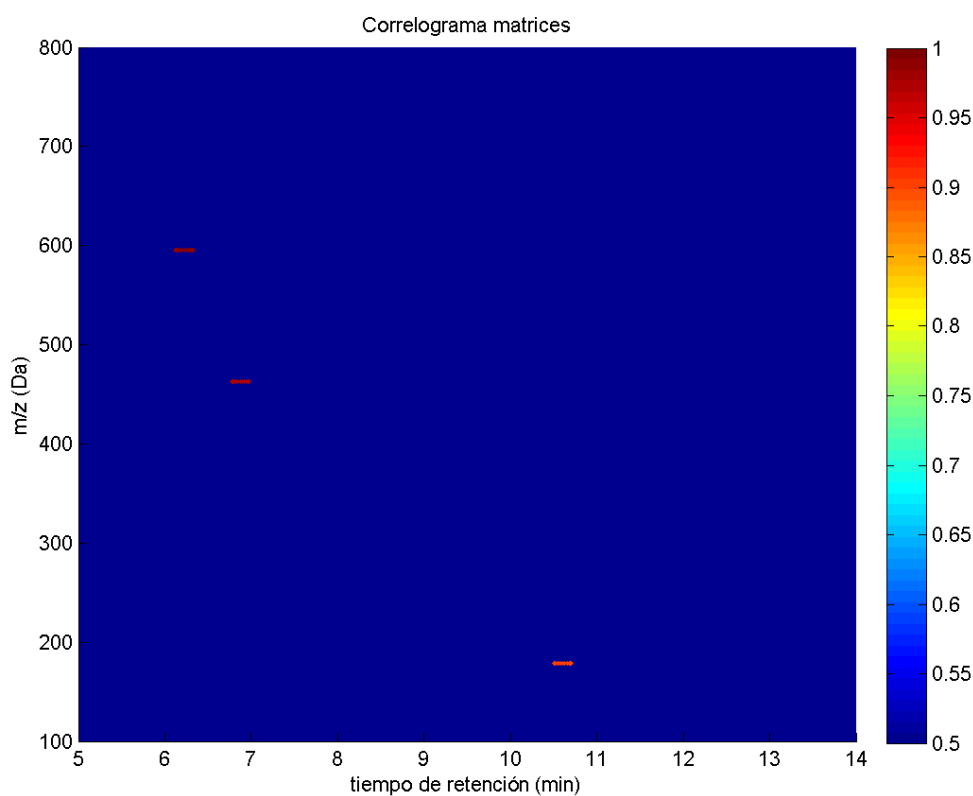


Figura 68. Representación de los coeficientes de correlación ($r > 0.5$) obtenidos utilizando la correlación entre el área de los compuestos obtenidos con *clustering* de líneas y la actividad de las 20 muestras de entrenamiento del experimento.



- **Predicción de la actividad.**

El modelo de predicción se ha estimado para cada uno de los 3 compuestos candidatos mediante una regresión lineal entre el área cromatográfica A_n y las actividades y_n , usando las 20 muestras de entrenamiento. El análisis de regresión proporciona la pendiente, la ordenada en el origen, el coeficiente de correlación y la semi-anchura del intervalo de confianza del 95% de la predicción de actividad para cada candidato (Tabla 16). De este modo, con el modelo de predicción y usando la Ecuación 10 (del apartado F de la metodología), se puede estimar o predecir la actividad a partir del área cromatográfica de los picos seleccionados de muestras nuevas.

Para ello, se ha calculado el área cromatográfica de los 3 candidatos proporcionados por el análisis anterior (son los mismos con ambos métodos) y se ha obtenido el modelo de predicción mediante regresión entre A_n e y_n para cada uno de los candidatos.

El modelo de predicción se ha evaluado con las muestras en los grupos de entrenamiento y evaluación. En la Figura 69 se muestra la actividad en función del área cromatográfica del primero de los candidatos A_{n1} para ambos grupos (círculos blancos para entrenamiento y círculos negros para evaluación). Se indican también la línea de regresión y los límites del intervalo de confianza del 95% de la actividad. La actividad predicha corresponde a la regresión lineal evaluada en el área medida, y el error de predicción se puede calcular como la diferencia entre la actividad medida y la actividad predicha (es decir, la distancia entre los círculos y la línea de regresión). Como puede verse, la línea de regresión describe aproximadamente la relación entre el área del pico cromatográfico y la actividad para ambos grupos.



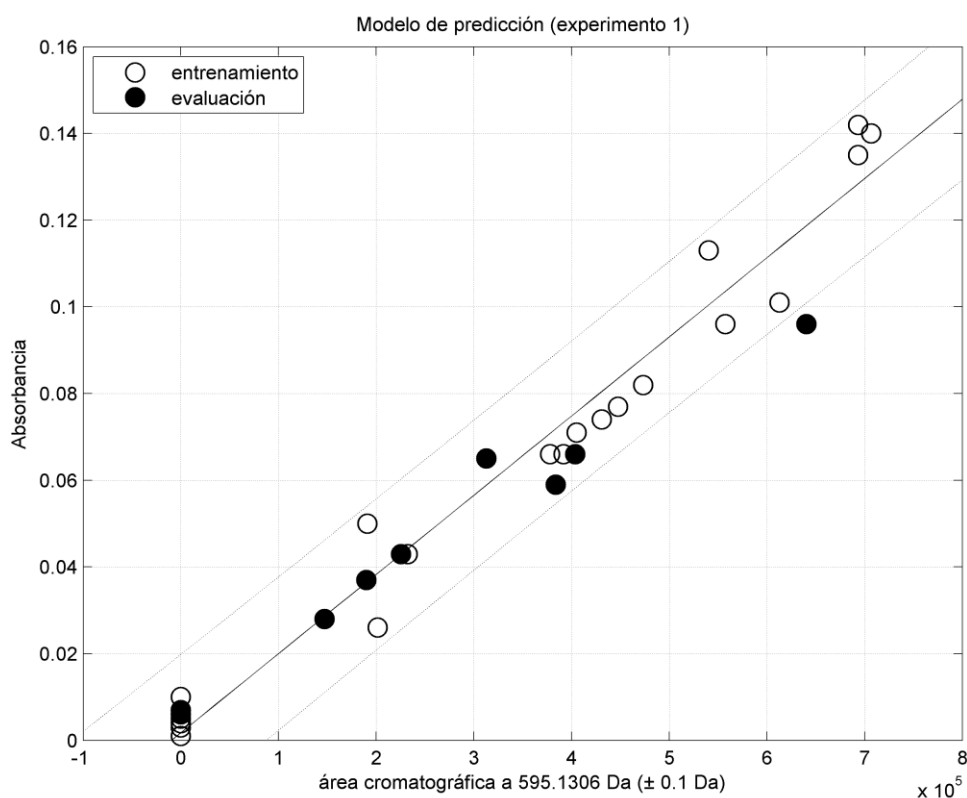


Figura 69. Actividad (absorbancia) en función del área cromatográfica A_n para el primer candidato ($m/z=595.1306$ Da; $t_R=6.25$ min). Los círculos blancos representan las muestras de entrenamiento; los círculos negros corresponden a las muestras de evaluación. Las líneas sólida es la recta de regresión mientras que la línea punteada representa los límites del intervalo de confianza del 95% (estimado con las muestras de entrenamiento).

Cada candidato a compuesto activo se ha evaluado comparando los coeficientes de correlación obtenidos con los grupos de entrenamiento y evaluación. Además, la exactitud del procedimiento de predicción se puede evaluar con el error cuadrático medio (rmse) entre las medidas y las predicciones de la actividad. La Tabla 16 muestra, para cada candidato, los principales resultados del análisis de regresión entre las áreas y las actividades. La tabla muestra el coeficiente de correlación obtenido de las muestras de entrenamiento y de las de evaluación. Como se puede apreciar, el coeficiente de correlación para los candidatos 1 y 2 es mayor que el del candidato 3. Además, los coeficientes de



correlación de las muestras de entrenamiento y evaluación son muy parecidos en los candidatos 1 y 2, mientras que el coeficiente de correlación de las muestras de evaluación es significativamente menor para el candidato 3. Con este resultado, el candidato 3 puede descartarse como candidato a compuesto activo. Como se discutió anteriormente, la correlación relativamente alta observada para este candidato está asociada con la selección de la concentración cuando las muestras fueron preparadas en el laboratorio y no a la actividad de este compuesto. Los rmse relativos a la desviación estándar de la actividad para ambos, entrenamiento y evaluación, también se muestran en la tabla. Los errores son bajos para los candidatos 1 y 2, pero significativamente grandes para el candidato 3. La tabla también incluye, para cada candidato, el valor de la ordenada en el origen, la pendiente, la correspondiente desviación estándar y la semi-anchura del intervalo de confianza del 95% de las estimaciones de la actividad basadas en la medida de las áreas cromatográficas.

Tabla 16. Resultados del análisis de regresión entre el área cromatográfica y la actividad para los 3 candidatos. Se comparan los coeficientes de correlación r obtenidos para las muestras de entrenamiento y evaluación. Para cada candidato se presenta, la desviación estándar relativa de la actividad (rel. rmse), la intersección con el eje y la pendiente (y su correspondiente desviación estándar) y la semi-anchura del intervalo de confianza del 95%.

	candidato 1	candidato 2	candidato 3
m/z (Da)	595.1306	463.0864	179.0350
t_R (min)	6.25	6.9	10.6
r (entrenamiento)	0.9840	0.9802	0.9046
r (evaluación)	0.9881	0.9900	0.6807
rel. rmse (entren.)	17.34 %	19.31 %	41.54 %
rel. rmse (eval.)	19.85 %	17.32%	45.77 %
ordenada origen	1.7444×10^{-3}	2.461×10^{-3}	1.629×10^{-2}
SE (ordenada origen)	3.320×10^{-3}	3.679×10^{-3}	7.1282×10^{-3}
Pendiente	1.8264×10^{-7}	1.0159×10^{-6}	1.2496×10^{-7}
SE (pendiente)	7.7834×10^{-9}	4.8401×10^{-8}	1.3879×10^{-8}
s.w. int. conf. 95%	0.01791	0.01996	0.04311

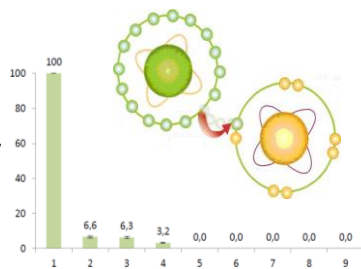


Teniendo en cuenta los resultados de la Tabla 16, se puede concluir que, con una alta probabilidad, el candidato 3 (ácido caféico) no es el compuesto activo, y uno de los otros candidatos (candidato 1, delfinidina-3-O-sambubioside o candidato 2, desconocido) es el compuesto activo. Probablemente el candidato 1 es el activo principal (porque su área es 5.6 veces mayor que la del candidato 2), y de hecho, en este experimento controlado, sabemos que el candidato 1 es el responsable de la actividad y el candidato 2 es un compuesto no activo que tiene una concentración fuertemente correlacionada con la del candidato 1. Sin embargo, si no hubiera conocimiento a priori sobre él, no podríamos concluir cuál de estos dos candidatos es el activo. En cualquier caso, la predicción obtenida de ambos candidatos proporciona una correlación alta y se podría concluir que uno de ellos es el compuesto activo y el otro es un compuesto cuya concentración está estadísticamente asociada a la del otro (probablemente contaminante o descomposición).





RESULTADOS Y DISCUSIÓN



BLOQUE I

Caracterización e identificación de compuestos antioxidantes en muestras de frutas y vegetales

BLOQUE 1. CARACTERIZACIÓN E IDENTIFICACIÓN DE COMPUESTOS ANTIOXIDANTES EN MUESTRAS DE FRUTAS Y VEGETALES.

El consumo de frutas y vegetales se ha asociado con la protección frente a enfermedades causadas por el estrés oxidativo como pueden ser las enfermedades cardiovasculares, el cáncer y las patologías neurodegenerativas^{140,141}. Estos efectos beneficiosos derivados del consumo de frutas y vegetales se han atribuido en parte a los componentes antioxidantes de las mismas¹⁴². Para saber específicamente qué compuestos procedentes de las frutas y vegetales son los responsables de la capacidad antioxidante, se precisa del estudio exhaustivo de cada uno de sus componentes. Como paso previo al estudio pormenorizado de los constituyentes antioxidantes, es necesario realizar un cribado con un amplio número de muestras, basado en la evaluación de la bioactividad de las mismas¹⁴³. Por ello, el primer objetivo del presente bloque es determinar la capacidad antioxidante de una gran variedad de frutas y vegetales del sur de España, principalmente Andalucía, para aportar una visión general de la capacidad antioxidante que proporcionan al consumirlas. El siguiente objetivo es seleccionar, en base a los resultados de este amplio estudio de capacidad antioxidante, una de las muestras para realizar un estudio exhaustivo de sus constituyentes. El fin último del estudio presentado en este bloque es estudiar la relación entre la composición química y la capacidad antioxidante de muestras vegetales. La Figura 70 muestra de

¹⁴⁰ Fu, L. *et al.*, 2011. Antioxidant capacities and total phenolic contents of 62 fruits. *Food Chemistry*, 129(2), pp.345–350.

¹⁴¹ Hervert-Hernández, D. *et al.*, 2011. The contribution of fruits and vegetables to dietary intake of polyphenols and antioxidant capacity in a Mexican rural diet: Importance of fruit and vegetable variety. *Food Research International*, 44(5), pp.1182–1189.

¹⁴² Podsedek, A., 2007. Natural antioxidants and antioxidant capacity of *Brassica* vegetables: A review. *LWT - Food Science and Technology*, 40(1), pp.1–11.

¹⁴³ Guo, C. *et al.*, 2003. Antioxidant activities of peel, pulp and seed fractions of common fruits as determined by FRAP assay. *Nutrition Research*, 23(12), pp.1719–1726.



manera esquemática los estudios realizados en este bloque, siguiendo el modelo del esquema de la metodología general de la tesis.

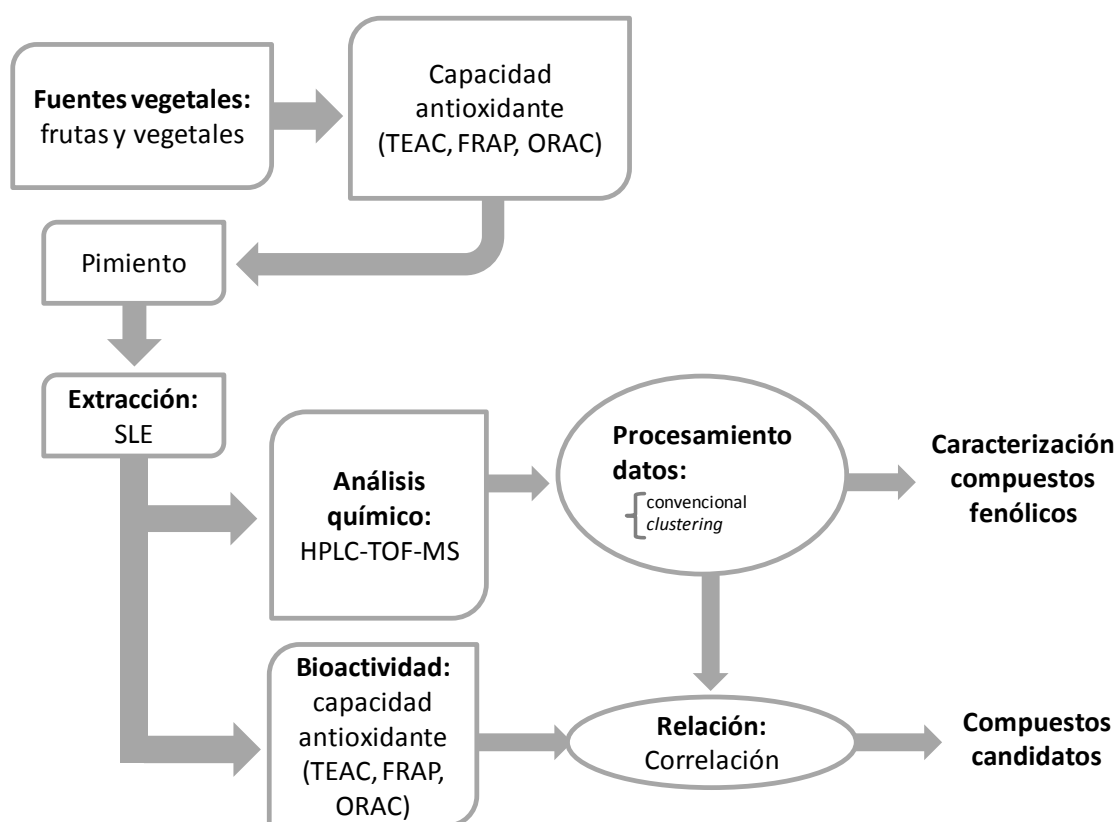


Figura 70. Esquema de los estudios realizados en el bloque 1.

A continuación, se presentan y discuten los resultados obtenidos en este bloque, divididos en tres apartados referentes al estudio de la capacidad antioxidante, el estudio de caracterización de las muestras de pimiento y el estudio de correlación, respectivamente.

1.1. Evaluación de la capacidad antioxidante de cuarenta y cuatro variedades de frutas y vegetales cultivados en el sur de España.

Se seleccionaron muestras de aguacate, caqui, chirimoya, granada, limón, mango, melón, membrillo, níspero, sandía, uva, ajo, alcachofa, berenjena, calabacín, cebolla, espárrago, haba, judía verde, lechuga, patata, pimiento,

tomate y zanahoria. De cada especie, se seleccionó a su vez entre una y tres variedades, obteniendo un total de 44 muestras. En la Figura 22 del apartado A.1 de la metodología se pueden consultar las variedades seleccionadas, así como la estacionalidad de las mismas. Todas las muestras se recolectaron a tiempos diferentes en el mismo año, dentro del correspondiente periodo de cosecha de cada vegetal (de dos a cinco tiempos diferentes de adquisición). Lo que se pretende es ofrecer al consumidor una idea de las propiedades beneficiosas de estas especies vegetales disponibles en el mercado a lo largo del año.

Para la evaluación de la capacidad antioxidante, se utilizaron tres métodos diferentes de capacidad antioxidante: TEAC, FRAP y ORAC. Este hecho, junto con el de adquirir las muestras a diferentes tiempos, implicó la realización de 154 análisis para las frutas y 339 análisis en el caso de los vegetales. Teniendo en cuenta además que los análisis se realizaron por triplicado, el total de análisis realizados en este trabajo asciende a 1479.

La capacidad antioxidante de todas las variedades de frutas y vegetales analizados por los tres métodos se resume en la Tabla 17 y la Tabla 18, respectivamente. Los datos se presentan como la media \pm desviación estándar (SD) para todas las muestras recolectadas a diferentes tiempos. Ambas tablas también incluyen los porcentajes de humedad obtenidos antes del proceso de liofilización. Los resultados aparecen calculados por 100 g de muestra (peso seco) y expresados como mmol equivalentes (eq.) de Trolox para los ensayos TEAC y ORAC, y mmol eq. de sulfato férrico (FeSO_4) en el caso del ensayo FRAP (Figura 71, Figura 72 y Figura 73 para TEAC, FRAP y ORAC, respectivamente). Estas gráficas muestran los resultados como un rango de capacidad antioxidante relacionado con el periodo de disponibilidad de las muestras en el mercado.



Tabla 17. Capacidad antioxidante (métodos TEAC, FRAP y ORAC) por 100 g (peso seco) y % de humedad para las muestras de frutas.

Frutas	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
AGUACATE					
Aguacate "Lamb-Hass" 1	Marzo	1.58±0.05	1.3±0.1	1.63±0.03	76
Aguacate "Lamb-Hass" 2	Junio	2.1±0.1	1.090±0.009	2.14±0.02	
Aguacate "Bacon" 1	Junio	3.1±0.1	2.02±0.06	0.81±0.03	77
Aguacate "Bacon" 2	Octubre	1.281±0.007	1.33±0.06	0.20±0.01	
CAQUI					
Caqui "Persimón" 1	Marzo	0.63±0.03	0.71±0.01	--	
Caqui "Persimón" 2	Octubre	0.41±0.02	0.730±0.009	0.31±0.01	86
Caqui "Persimón" 3	Noviembre	0.33±0.02	0.660±0.004	0.25±0.02	
CHIRIMOYA					
Chirimoya "Fino de Jete" 1	Enero	2.0±0.2	1.7±0.1	0.83±0.02	
Chirimoya "Fino de Jete" 2	Febrero	2.3±0.1	2.9±0.1	0.97±0.02	95
Chirimoya "Fino de Jete" 3	Marzo	2.5±0.1	2.597±0.007	1.31±0.03	
Chirimoya "Fino de Jete" 4	Marzo	2.3±0.2	2.63±0.05	1.14±0.08	
GRANADA					
Granada "Mollar de Elche" 1	Agosto	3.8±0.4	4.83±0.08	1.344±0.008	
Granada "Mollar de Elche" 2	Septiembre	5.8±0.2	5.4±0.2	1.10±0.07	93
Granada "Mollar de Elche" 3	Octubre	5.5±0.2	5.7±0.2	0.91±0.03	



Frutas	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
LIMÓN					
Limón "Fino" 1	Septiembre	5.3±0.2	14.3±0.6	2.91±0.07	
Limón "Fino" 2	Noviembre	5.9±0.4	11.6±0.7	3.7±0.1	90
Limón "Fino" 3	Diciembre	3.3±0.2	12.5±0.9	2.92±0.05	
Limón "Primaflori" 1	Mayo	4.70±0.07	13.3±0.1	2.08±0.02	
Limón "Primaflori" 2	Agosto	4.8±0.4	12.8±0.2	--	91
Limón "Primaflori" 3	Octubre	5.4±0.2	12.07±0.09	2.4±0.1	
Limón "Primaflori" 4	Noviembre	4.0±0.2	11.55±0.10	2.39±0.03	
MANGO					
Mango "Sensación" 1	Agosto	3.9±0.1	4.06±0.08	0.45±0.02	
Mango "Sensación" 2	Septiembre	4.2±0.2	4.17±0.04	0.69±0.03	85
Mango "Sensación" 3	Septiembre	3.6±0.2	3.82±0.07	0.68±0.01	
Mango "Sensación" 4	Octubre	3.2±0.3	3.5±0.1	0.69±0.02	
MELÓN					
Melón "Cantaluz" 1	Junio	2.7±0.1	3.34±0.03	0.47±0.03	
Melón "Cantaluz" 2	Julio	0.94±0.09	2.29±0.07	0.42±0.02	93
Melón "Cantaluz" 3	Julio	1.57±0.05	2.06±0.03	0.42±0.03	
Melón "Cantaluz" 4	Octubre	1.35±0.06	1.23±0.04	0.48±0.01	



Frutas	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
Melón "Galia" 1	Junio	1.2±0.1	1.96±0.04	0.28±0.02	
Melón "Galia" 2	Junio	1.2±0.1	1.91±0.07	0.22±0.01	90
Melón "Galia" 3	Julio	1.6±0.1	2.12±0.06	--	
Melón "Galia" 4	Octubre	1.03±0.09	2.38±0.03	0.53±0.03	
Melón "Piel de Sapo" 1	Junio	1.04±0.08	2.17±0.03	0.19±0.01	
Melón "Piel de Sapo" 2	Julio	1.17±0.07	1.91±0.02	0.40±0.03	90
Melón "Piel de Sapo" 3	Julio	0.91±0.03	1.73±0.03	0.42±0.02	
Melón "Piel de Sapo" 4	Octubre	1.00±0.09	1.78±0.03	0.398±0.002	
MEMBRILLO					
Membrillo "Lusitanica" 1	Julio	0.77±0.08	1.62±0.09	0.46±0.01	
Membrillo "Lusitanica" 2	Agosto	0.71±0.03	1.26±0.04	0.344±0.003	85
Membrillo "Lusitanica" 3	Octubre	1.17±0.09	2.41±0.07	0.581±0.005	
Membrillo "Lusitanica" 4	Noviembre	0.28±0.04	0.92±0.03	0.39±0.01	
NÍSPERO					
Níspero "Algerie" 1	Mayo	0.23±0.05	0.56±0.03	---	
Níspero "Algerie" 2	Mayo	0.61±0.06	0.66±0.02	0.38±0.02	84
Níspero "Algerie" 3	Junio	0.6±0.1	0.59±0.02	0.610±0.002	



Frutas	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
SANDÍA					
Sandía "Fashion" 1	Junio	0.33±0.01	0.62±0.02	--	
Sandía "Fashion" 2	Julio	0.69±0.07	0.86±0.03	0.55±0.01	
Sandía "Fashion" 3	Julio	0.586±0.002	0.71±0.01	0.54±0.02	92
Sandía "Fashion" 4	Agosto	0.64±0.06	0.67±0.01	0.36±0.02	
Sandía "Fashion" 5	Agosto	1.13±0.10	0.82±0.03	0.433±0.003	
UVA					
Uva roja 1	Agosto	4.0±0.1	13.2±0.2	2.42±0.05	
Uva roja 2	Septiembre	3.2±0.2	16.4±0.3	3.58±0.08	79
Uva roja 3	Noviembre	4.4±0.3	14.0±0.2	2.7±0.1	
Uva roja 4	Diciembre	4.6±0.7	13.8±0.4	2.7±0.1	

"--" para valores no disponibles.



Tabla 18. Capacidad antioxidante (métodos TEAC, FRAP y ORAC) por 100 g (peso seco) y % de humedad para las muestras de vegetales.

Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
AJO					
Ajo blanco 1	Mayo	0.13±0.07	0.39±0.02	2.34±0.05	
Ajo blanco 2	Agosto	0.6±0.2	0.58±0.06	4.3±0.3	67
Ajo blanco 3	Septiembre	0.3±0.2	0.54±0.02	1.96±0.06	
Ajo blanco 4	Octubre	0.24±0.03	0.47±0.03	0.83±0.05	
Ajo rojo 1	Junio	0.36±0.03	0.83±0.03	--	
Ajo rojo 2	Agosto	0.59±0.03	0.71±0.02	3.8±0.1	69
Ajo rojo 3	Septiembre	0.40±0.01	0.72±0.02	3.6±0.2	
Ajo rojo 4	Octubre	0.28±0.02	0.55±0.01	3.7±0.5	
ALCACHOFA					
Alcachofa "Blanca de Tudela" 1	Febrero	0.67±0.03	1.10±0.03	0.853±0.002	
Alcachofa "Blanca de Tudela" 2	Marzo	0.79±0.04	1.08±0.03	0.45±0.03	84
Alcachofa "Blanca de Tudela" 3	Mayo	1.45±0.06	2.9±0.1	0.81±0.08	
Alcachofa "Blanca de Tudela" 4	Noviembre	1.003±0.008	1.72±0.02	0.99±0.05	
BERENJENA					
Berenjena púrpura rayada 1	Marzo	3.5±0.3	6.432±0.04	--	
Berenjena púrpura rayada 2	Abril	2.9±0.2	6.34±0.08	2.09±0.09	
Berenjena púrpura rayada 3	Mayo	3.2±0.1	5.1±0.1	0.80±0.02	93
Berenjena púrpura rayada 4	Octubre	3.0±0.2	6.6±0.2	0.80±0.05	
Berenjena púrpura rayada 5	Diciembre	2.78±0.04	5.6±0.2	1.01±0.02	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
Berenjena redonda 1	Febrero	1.420±0.002	2.30±0.09	--	
Berenjena redonda 2	Abril	1.20±0.01	2.22±0.07	1.28±0.09	92
Berenjena redonda 3	Septiembre	1.32±0.08	1.894±0.007	0.60±0.03	
Berenjena larga 1	Febrero	4.29±0.05	8.0±0.5	--	
Berenjena larga 2	Marzo	3.6±0.1	8.7±0.2	--	
Berenjena larga 3	Septiembre	5.14±0.06	5.3±0.2	0.38±0.03	93
Berenjena larga 4	Octubre	4.9±0.4	7.6±0.4	0.519±0.002	
Berenjena larga 5	Octubre	4.3±0.2	7.8±0.5	0.59±0.01	
CALABACÍN					
Calabacín verde 1	Febrero	2.2±0.2	1.780±0.007	1.42±0.05	
Calabacín verde 2	Marzo	2.52±0.07	0.80±0.02	1.072±0.009	80
Calabacín verde 3	Junio	2.4±0.2	0.83±0.01	1.08±0.03	
Calabacín verde 4	Septiembre	1.03±0.06	0.78±0.02	1.02±0.02	
Calabacín redondo 1	Abril	1.09±0.07	0.66±0.01	0.61±0.02	
Calabacín redondo 2	Mayo	1.00±0.03	0.31±0.03	0.90±0.03	80
Calabacín redondo 3	Julio	0.44±0.03	0.36±0.04	0.48±0.03	
Calabacín redondo 4	Septiembre	0.578±0.004	0.54±0.06	0.499±0.006	
Calabacín ecológico 1	Marzo	1.64±0.10	3.09±0.08	1.728±0.002	
Calabacín ecológico 2	Junio	1.8±0.2	2.76±0.03	1.48±0.06	80
Calabacín ecológico 3	Septiembre	1.9±0.1	2.92±0.08	1.39±0.02	
Calabacín ecológico 4	Diciembre	2.16±0.05	2.66±0.05	1.66±0.06	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
CEBOLLA					
Cebolla blanca 1	Abril	3.19±0.07	3.4±0.2	1.60±0.02	
Cebolla blanca 2	Mayo	2.45±0.09	6.86±0.08	1.66±0.02	90
Cebolla blanca 3	Julio	2.86±0.23	3.25±0.20	2.06±0.03	
Cebolla blanca 4	Septiembre	5.6±0.2	3.3±0.2	0.52±0.01	
Cebolla roja 1	Enero	5.6±0.2	6.2±0.1	3.83±0.09	
Cebolla roja 2	Marzo	5.03±0.07	5.6±0.4	2.9±0.2	91
Cebolla roja 3	Agosto	5.126±0.009	5.7±0.3	3.8±0.1	
Cebolla roja 4	Octubre	4.0±0.4	5.1±0.2	3.83±0.03	
Cebolla tierna 1	Febrero	3.1±0.2	3.1±0.2	1.60±0.02	
Cebolla tierna 2	Marzo	2.4±0.1	2.4±0.3	1.66±0.02	89
Cebolla tierna 3	Septiembre	2.5±0.4	3.2±0.5	1.68±0.03	
Cebolla tierna 4	Noviembre	2.9±0.1	3.3±0.6	1.58±0.01	
ESPÁRRAGO					
Espárrago verde 1	Febrero	3.40±0.06	5.1±0.2	--	
Espárrago verde 2	Febrero	3.02±0.06	3.9±0.3	2.99±0.06	
Espárrago verde 3	Marzo	3.1±0.3	4.5±0.4	3.2±0.2	94
Espárrago verde 4	Marzo	3.7±0.3	4.8±0.2	2.88±0.03	
Espárrago verde 5	Marzo	3.9±0.1	4.8±0.2	3.0±0.2	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
HABA					
Haba verde 1	Enero	5.4±0.2	3.95±0.04	1.29±0.03	
Haba verde 2	Junio	4.8±0.1	4.3±0.1	2.53±0.02	62
Haba verde 3	Diciembre	4.7±0.1	4.06±0.09	2.25±0.08	
Haba verde 4	Diciembre	4.05±0.08	3.13±0.05	2.28±0.03	
JUDÍA VERDE					
Judía verde "Strike" 1	Abril	1.1±0.1	1.52±0.04	0.5±0.03	
Judía verde "Strike" 2	Septiembre	1.5±0.1	1.72±0.04	1.00±0.05	93
Judía verde "Perona" 1	Febrero	0.50±0.02	1.00±0.05	0.55±0.03	
Judía verde "Perona" 2	Agosto	0.59±0.02	0.93±0.02	0.86±0.05	
Judía verde "Perona" 3	Septiembre	0.27±0.04	0.48±0.03	0.600±0.003	93
Judía verde "Perona" 4	Octubre	0.49±0.05	0.650±0.008	0.67±0.03	
Judía verde "Helda " 1	Febrero	0.23±0.02	0.36±0.02	--	
Judía verde "Helda "2	Agosto	0.31±0.03	0.38±0.02	--	93
Judía verde "Helda "3	Septiembre	0.69±0.03	0.33±0.03	0.610±0.009	
Judía verde "Helda "4	Octubre	0.32±0.01	0.692±0.008	0.53±0.03	
LECHUGA					
Lechuga "Romana" 1	Enero	0.47±0.05	0.446±0.003	0.48±0.05	
Lechuga "Romana" 2	Febrero	0.28±0.05	0.427±0.009	0.52±0.02	95
Lechuga "Romana" 3	Mayo	0.93±0.07	1.77±0.05	0.97±0.02	
Lechuga "Romana" 4	Noviembre	1.02±0.03	1.97±0.05	0.87±0.01	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
Lechuga "Iceberg" 1	Enero	0.08±0.01	0.38±0.02	0.50±0.06	97
Lechuga "Iceberg" 2	Abril	0.40±0.03	0.523±0.001	0.63±0.05	
Lechuga "Iceberg" 3	Noviembre	1.67±0.02	2.1±0.1	0.62±0.02	
Lechuga "Baby" 1	Marzo	3.4±0.2	6.0±0.6	1.99±0.08	95
Lechuga "Baby" 2	Octubre	2.5±0.2	5.36±0.02	1.78±0.08	
Lechuga "Baby" 3	Diciembre	2.7±0.1	6.4±0.2	2.00±0.06	
PATATA					
Patata "Vivaldi" 1	Febrero	0.59±0.02	0.399±0.008	0.286±0.006	
Patata "Vivaldi" 2	Abril	0.42±0.02	0.20±0.02	0.2622±0.0006	77
Patata "Vivaldi" 3	Julio	0.04±0.01	0.22±0.01	0.238±0.003	
Patata "Vivaldi" 4	Diciembre	0.45±0.02	0.29±0.02	0.283±0.005	
Patata "Scarlet" 1	Enero	0.28±0.02	0.41±0.02	0.73±0.01	
Patata "Scarlet" 2	Febrero	0.270±0.004	0.52±0.01	0.75±0.03	77
Patata "Scarlet" 3	Marzo	0.59±0.02	0.680±0.006	0.73±0.03	
Patata "Scarlet" 4	Noviembre	0.43±0.03	0.70±0.01	0.550±0.008	
PIMIENTO					
Pimiento "Lamuyo" amarillo 1	Enero	11.2±0.6	13.8±0.6	2.77±0.09	
Pimiento "Lamuyo" amarillo 2	Agosto	10.3±0.6	10.9±0.3	1.65±0.03	89
Pimiento "Lamuyo" amarillo 3	Septiembre	5.0±0.5	5.5±0.3	1.248±0.005	
Pimiento "Lamuyo" amarillo 4	Octubre	9.4±0.5	8.7±0.2	2.43±0.08	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
Pimiento "California" rojo 1	Febrero	4.6±0.2	8.0±0.2	0.33±0.02	
Pimiento "California" rojo 2	Marzo	5.3±0.1	7.2±0.1	1.478±0.006	89
Pimiento "California" rojo 3	Agosto	10.8±0.2	12.1±0.2	1.477±0.005	
Pimiento "California" rojo 4	Septiembre	13.8±0.2	11.3±0.4	1.50±0.02	
Pimiento Italiano verde 1	Febrero	4.6±0.3	5.0±0.7	--	
Pimiento Italiano verde 2	Marzo	5.310±0.004	6.12±0.05	1.46±0.02	
Pimiento Italiano verde 3	Agosto	3.8±0.4	4.1±0.1	1.57±0.02	92
Pimiento Italiano verde 4	Septiembre	3.4±0.1	5.9±0.2	1.88±0.08	
Pimiento Italiano verde 5	Septiembre	4.1±0.2	3.8±0.1	2.29±0.05	
TOMATE					
Tomate verde "Raf" 1	Enero	1.5±0.1	6.0±0.2	--	
Tomate verde "Raf" 2	Marzo	1.48±0.08	5.47±0.08	0.97±0.03	
Tomate verde "Raf" 3	Mayo	1.38±0.08	6.00±0.07	0.83±0.06	93
Tomate verde "Raf" 4	Julio	1.8±0.1	6.7±0.6	1.22±0.03	
Tomate verde "Raf" 5	Octubre	2.0±0.1	4.6±0.5	1.34±0.03	



Vegetales	Fecha recolección (2011)	TEAC (mmol eq. Trolox/100 g)	FRAP (mmol eq. FeSO ₄ /100 g)	ORAC (mmol eq. Trolox/100 g)	% Humedad
Tomate rojo "Daniela" 1	Febrero	1.7±0.2	3.03±0.06	0.84±0.02	
Tomate rojo "Daniela" 2	Mayo	2.0±0.1	3.56±0.05	1.02±0.05	93
Tomate rojo "Daniela" 3	Agosto	1.8±0.1	5.0±0.2	1.25±0.03	
Tomate rojo "Daniela" 4	Octubre	1.3±0.2	4.6±0.2	1.19±0.03	
ZANAHORIA					
Zanahoria "Nangro"1	Mayo	1.48±0.07	1.43±0.01	0.63±0.01	
Zanahoria "Nangro" 2	Junio	0.44±0.03	0.390±0.002	--	91
Zanahoria "Nangro"3	Septiembre	0.53±0.05	1.44±0.03	0.369±0.008	
Zanahoria "Nangro"4	Noviembre	0.17±0.05	0.54±0.05	0.354±0.008	

--" para valores no disponibles.



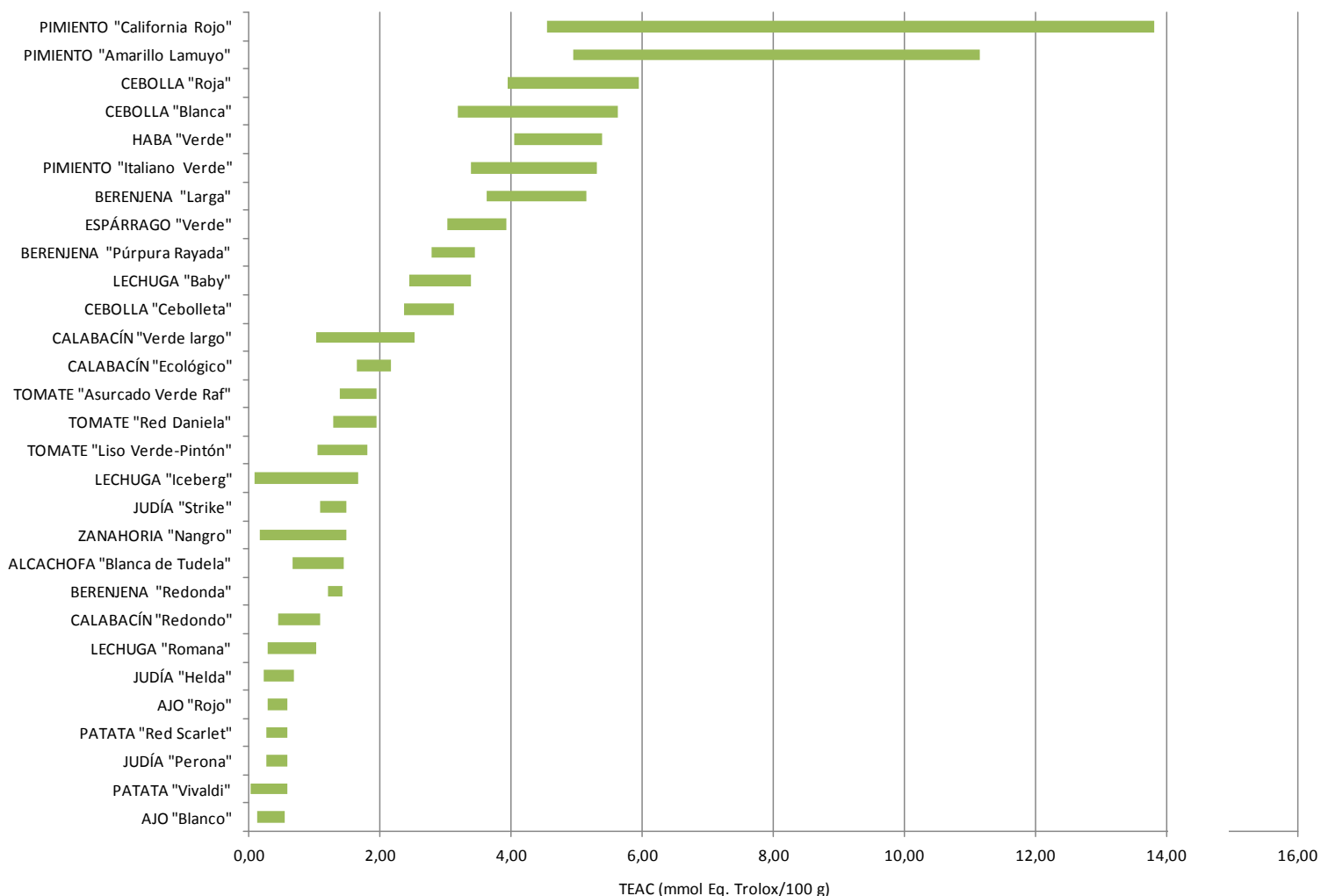
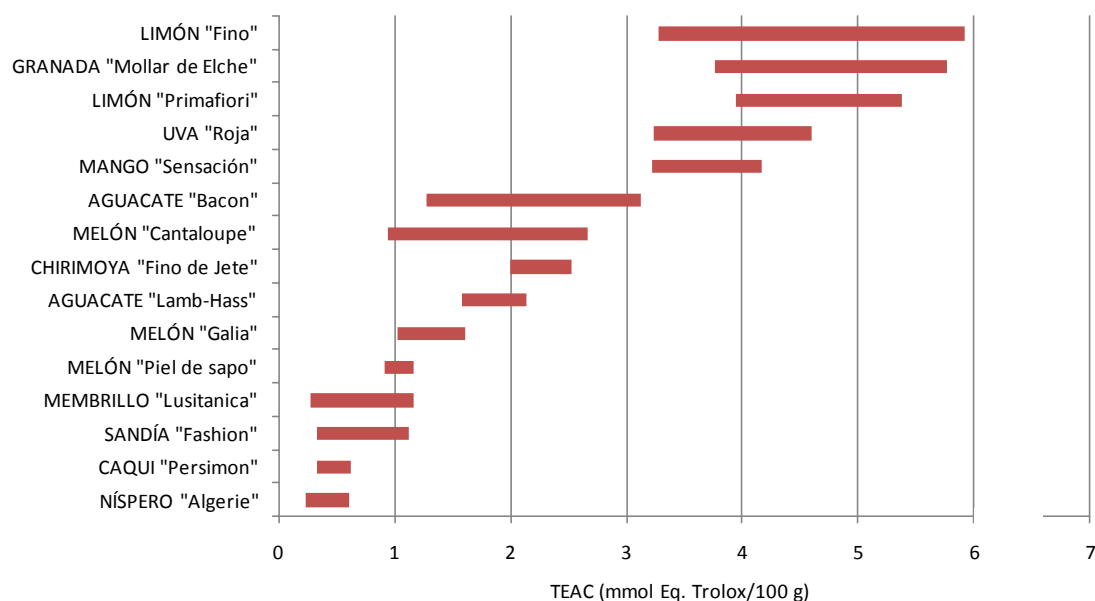


Figura 71. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método TEAC.



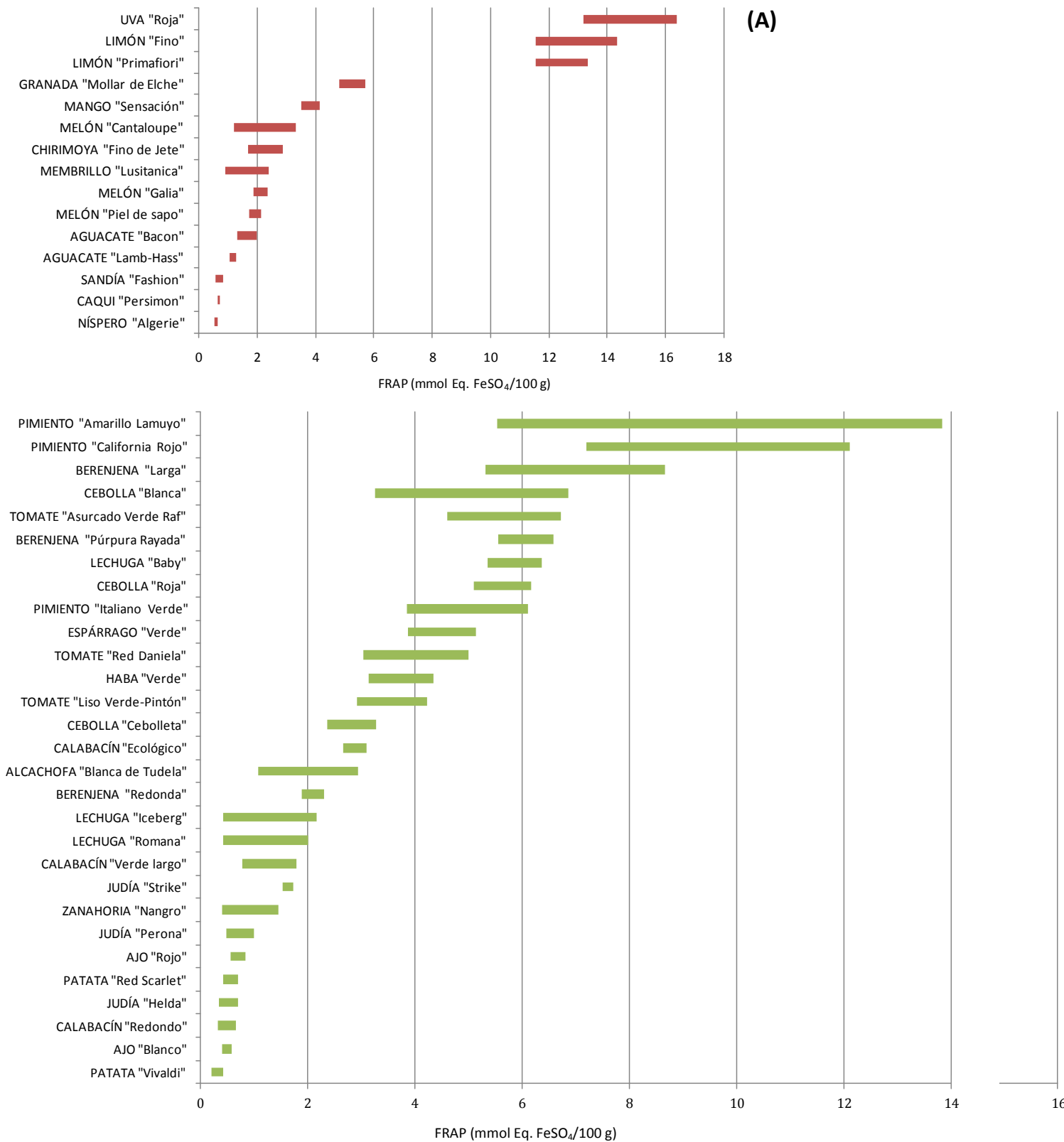


Figura 72. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método FRAP.



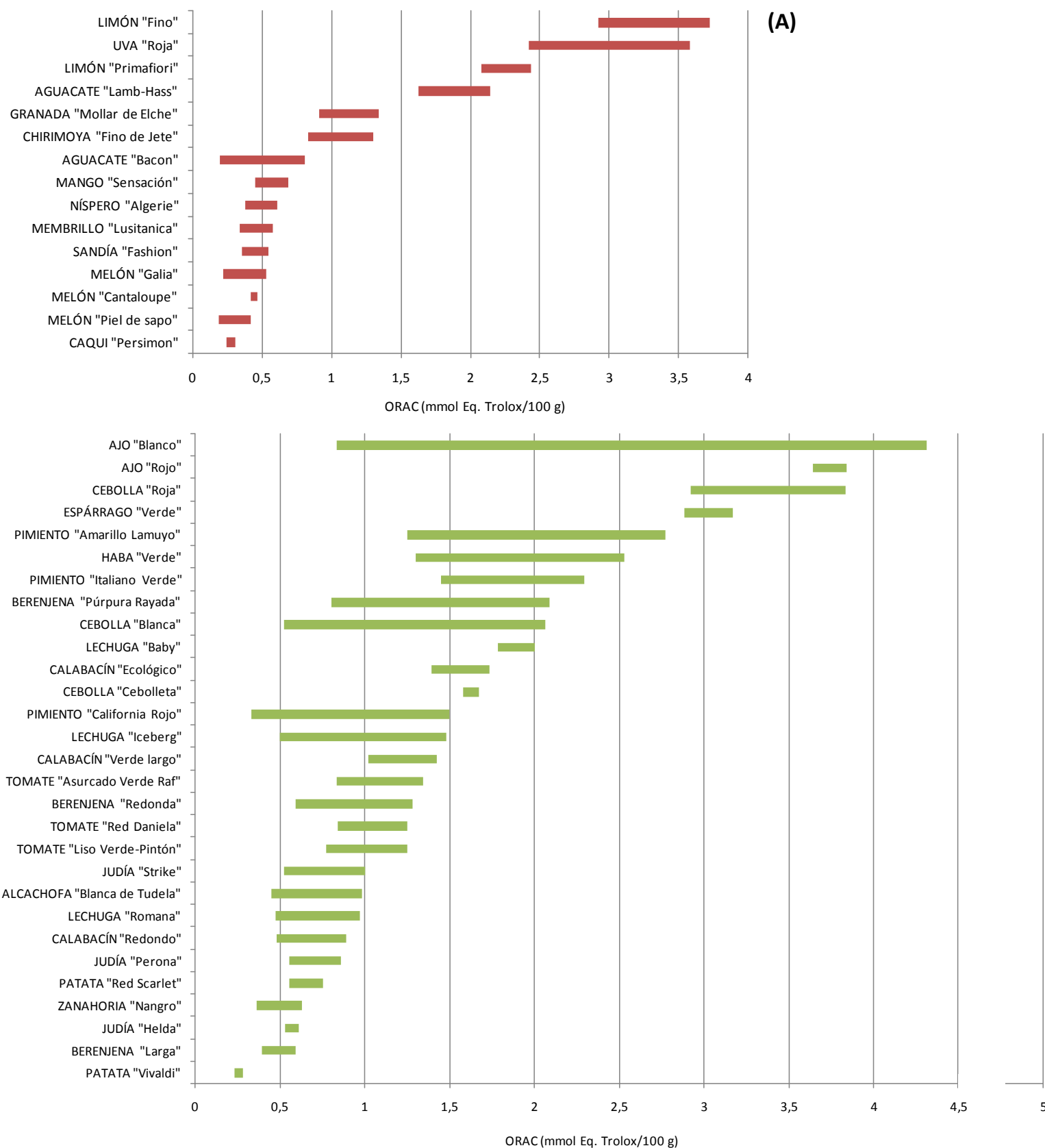


Figura 73. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método ORAC.



Cabe destacar que los resultados obtenidos variaron dependiendo del método utilizado; es decir, los valores de capacidad antioxidante obtenidos mediante los métodos basados en la transferencia de electrones (ensayos TEAC y FRAP) difirieron de los resultados obtenidos con el método basado en la transferencia de átomos de hidrógeno (ensayo ORAC).

Generalmente, en los casos en los que la porción comestible está constituida únicamente por la pulpa, los valores de capacidad antioxidante fueron menores que los resultados de otros estudios que incluyen piel y/o semillas. De hecho, se ha estudiado ampliamente la capacidad antioxidante de subproductos procedentes de industrias tales como la vitivinícola, cuyos residuos han demostrado ser excelentes fuentes de antioxidantes¹⁴⁴.

Algunos estudios han revelado que el genotipo es el parámetro de mayor influencia sobre la composición fitoquímica de las frutas y vegetales y, por tanto, sobre su capacidad antioxidante. Por esta razón, los resultados de la capacidad antioxidante de diferentes variedades de la misma especie son similares¹⁴⁵. En este estudio, la capacidad antioxidante varió en función del genotipo y del método utilizado; también estuvo influenciada por el estado fisiológico de las especies vegetales. Como puede verse en la Figura 71, Figura 72 y Figura 73, se obtuvieron rangos de capacidad antioxidante muy amplios para algunas variedades como es el caso de el pimiento rojo y el amarillo y el ajo blanco. Esto reveló que la capacidad antioxidante varía significativamente entre aquellas muestras de la misma variedad recolectadas a diferentes tiempos. Algunos ejemplos de estas diferencias se explican a continuación, comparando

¹⁴⁴ Arvanitoyannis, I.S., Ladas, D. & Mavromatis, A., 2006. Potential uses and applications of treated wine waste: a review. *International Journal of Food Science and Technology*, 41(5), pp.475–487.

¹⁴⁵ Rodríguez-Pérez, C. *et al.*, 2013. Comparative characterization of phenolic and other polar compounds in Spanish melon cultivars by using high-performance liquid chromatography coupled to electrospray ionization quadrupole-time of flight mass spectrometry. *Food Research International*, 54(2), pp.1519–1527.



los resultados con datos del Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA) sobre la temporalidad de las frutas y vegetales. Los resultados se explican por separado, en primer lugar, los resultados obtenidos para las frutas y en segundo lugar, los resultados derivados de los ensayos realizados con los vegetales.

Para las frutas, los valores TEAC variaron de 0.23 a 5.9 mmol eq. Trolox/100 g, los valores FRAP variaron desde 0.56 a 16.4 mmol eq. FeSO₄/100 g y los valores ORAC variaron desde 0.19 a 3.7 mmol eq. Trolox/100g. Generalmente, los valores ORAC fueron más bajos que los obtenidos en los otros dos ensayos.

Como se muestra en la Figura 71A, Figura 72A y en la Figura 73A, el limón, la uva roja y la granada fueron de las frutas que mostraron la capacidad antioxidante más alta en todos los métodos evaluados. La capacidad antioxidante de las dos variedades de limón fue similar según los valores obtenidos para TEAC y ORAC (Figura 71A y Figura 73A, respectivamente), mientras que en el ensayo FRAP los resultados fueron mucho mayores (Figura 72A). En todos los métodos evaluados, la capacidad antioxidante para el limón "Fino" fue ligeramente más alta que para el limón "Primafiori".

Para la granada y la uva roja, los estudios de actividad antioxidante son limitados, basándose la mayoría de las investigaciones en el estudio de extractos de zumos de estas frutas^{146,147}.

Los valores FRAP y TEAC para la variedad de granada "Mollar de Elche" variaron de 4.83 a 5.7 mmol eq. FeSO₄/100 g y de 3.8 a 5.8 mmol eq. Trolox/100g, respectivamente. Con respecto a las diferencias de la capacidad antioxidante

¹⁴⁶ Elfalleh, W. *et al.*, 2011. Antioxidant capacities of phenolic compounds and tocopherols from Tunisian pomegranate (*Punica granatum*) fruits. *Journal of Food Science*, 76(5), pp.C707–C713.

¹⁴⁷ El Kar, C. *et al.*, 2011. Pomegranate (*Punica granatum*) juices: chemical composition, micronutrient cations, and antioxidant capacity. *Journal of Food Science*, 76(6), pp.C795–C800.



entre las muestras de granada recolectadas a diferentes tiempos, se ha de destacar que los valores de TEAC y FRAP fueron similares para las muestras recolectadas en Septiembre y Octubre. Dichos valores fueron ligeramente menores para las muestras adquiridas en Agosto que para el resto (Tabla 1). Esto puede deberse a que Agosto corresponde con una época de recolección temprana de esta fruta¹⁴⁸. Fu *et al.* (2011) estudiaron un gran grupo de muestras mediante ensayos FRAP y TEAC, siendo sus resultados concordantes con los datos obtenidos en el presente trabajo en lo concerniente a la alta capacidad antioxidante de la granada¹⁴⁰. Sin embargo, los resultados ORAC no pueden compararse con los datos de bibliografía puesto que hay muy pocos estudios que analicen la porción comestible de granada y uva roja por este método.

Las muestras de fruta menos bioactivas fueron diferentes en función el método utilizado. En los ensayos TEAC y FRAP, las muestras que registraron la capacidad antioxidante más baja fueron el níspero, el caqui, y la sandía, en este orden. Sin embargo, las variedades de melón "Galia" y "Piel de sapo" junto con el aguacate "Bacon" presentaron la capacidad antioxidante más baja en el ensayo ORAC.

Los fenoles totales del níspero han sido ampliamente estudiados por otros autores¹⁴⁹. Sin embargo, es escasa la información disponible sobre la capacidad antioxidante de muestras de níspero analizadas mediante ORAC, TEAC o FRAP.

Otra muestra que manifestó una baja capacidad antioxidante en el ensayo TEAC fue el membrillo; esta información no puede contrastarse con datos de otros estudios porque la bibliografía disponible acerca de la capacidad antioxidante

¹⁴⁸ MAGRAMA (Ministerio de Agricultura, Alimentación y Medio Ambiente), 2013. Tabla informativa de frutas y verduras de temporada. Retrieved from http://www.alimentacion.es/es/campanas/frutas/frutas_verduras_temporada/.

¹⁴⁹ Gruz, J. *et al.*, 2011. Phenolic acid content and radical scavenging activity of extracts from medlar (*Mespilus germanica* L.) fruit at different stages of ripening. *Food Chemistry*, 124(1), pp.271–277.



de este fruto se refiere principalmente a mermeladas y/o procede de ensayos distintos como el DPPH^{150,151}.

En este estudio, el caqui registró valores TEAC de 0.33 a 0.63 mmol eq. Trolox/100 g y valores FRAP de 0.660 a 0.730 mmol eq. FeSO₄/100 g. Estos resultados concuerdan con la bibliografía. No hay datos disponibles para comparar los valores ORAC del caqui¹⁴³. La variedad analizada de sandía fue la denominada "Fashion", que proporcionó valores TEAC de 0.33 a 1.13 mmol eq. Trolox/100 g, valores FRAP de 0.62 a 0.86 mmol eq. FeSO₄/100 g y valores ORAC de 0.36 a 0.55 mmol eq. Trolox/100 g. Estos valores muestran una baja capacidad antioxidante de la sandía, lo que concuerda con los resultados de otros autores¹⁴⁰. Como puede observarse en la Tabla 1, los valores de capacidad antioxidante de las muestras de sandía recolectadas en Julio fueron parecidos entre sí para todos los métodos utilizados. El valor más bajo de TEAC se obtuvo con la primera muestra de sandía "Fashion". Esto puede deberse a que se recolectó en Junio, una fecha cercana a la fecha de recolección temprana¹⁴⁸.

De las variedades de melón "Cantaloupe", "Galia" y "Piel de sapo" analizadas en este estudio, la que mostró la capacidad antioxidante más baja fue la "Piel de sapo". Para esta muestra, los valores TEAC variaron de 0.91 a 1.17 mmol eq. Trolox/100 g, FRAP de 1.73 a 2.17 mmol eq. FeSO₄/100 g y ORAC de 0.19 a 0.42 mmol eq. Trolox/100 g. El melón es descrito por otros autores como una de las frutas con menor capacidad antioxidante. Los valores de baja capacidad antioxidante del melón y la sandía podrían deberse al hecho de que muchos de los compuestos que contribuyen a la capacidad antioxidante en estas frutas se encuentran en las semillas y la piel¹⁴³.

¹⁵⁰ Silva, B.M. *et al.*, 2004. Quince (*Cydonia oblonga* Miller) fruit (pulp, peel, and seed) and Jam: antioxidant activity. *Journal of Agricultural and Food Chemistry*, 52(15), pp.4705–4712.

¹⁵¹ Wojdyło, A. *et al.*, 2013. Composition and quantification of major polyphenolic compounds, antioxidant activity and colour properties of quince and mixed quince jams. *International Journal of Food Sciences and Nutrition*, 64(6), pp.749–756.



Para el aguacate "Bacon", los valores TEAC estuvieron en un rango de 1.281 a 3.1 mmol eq. Trolox/100 g, los valores FRAP estuvieron entre 1.33 y 2.02 mmol eq. FeSO₄/100 g y los valores ORAC fueron menores que en los otros métodos (de 0.20 a 0.81 mmol eq. Trolox/100 g). Con respecto a las diferencias de capacidad antioxidante entre los diferentes tiempos de adquisición de las muestras de aguacate, la variedad "Bacon" se recolectó en Marzo, Junio y Octubre. Los mejores resultados se obtuvieron con muestras las recolectadas en Junio para todos los métodos evaluados. En el caso de la variedad de aguacate "Lamb-Hass", los valores de capacidad antioxidante fueron mayores para las muestras recolectadas en Junio que para las de Marzo en los ensayos TEAC y ORAC. Estos resultados puede deberse a que el período de cosecha del aguacate es de Mayo a Noviembre, correspondiendo estos dos meses con los tiempos de recolección temprana y tardía, respectivamente¹⁴⁸.

A continuación, se da paso a la discusión de los resultados de los vegetales. Considerando la totalidad de vegetales analizados en el presente estudio, los valores TEAC estuvieron entre 0.04y 13.8 mmol eq. Trolox/100 g, los valores FRAC variaron de 0.20 a 13.8 mmol eq. FeSO₄/100 g y los valores ORAC fueron desde 0.238 hasta 4.3 mmol eq. Trolox/100 g.

Las variedades de pimiento rojo y amarillo mostraron la capacidad antioxidante más alta en ambos ensayos, TEAC y FRAP (Figura 71B y Figura 72B, respectivamente). De acuerdo con estos dos ensayos respectivamente, el tercer valor más alto correspondió con la cebolla roja y la berenjena larga. Bajorun *et al.* (2004) estudiaron un conjunto de 10 vegetales de Mauritania, cuyos resultados se dividieron en grupos de acuerdo al nivel de fenoles totales analizado. De esta manera, propusieron que el pimiento pertenecía a un grupo de nivel medio mientras que la cebolla sería de alto contenido en fenoles. En nuestro estudio, los valores TEAC y FRAP fueron al contrario, en el sentido de



que las muestras de pimiento mostraron mayor capacidad antioxidante que la cebolla blanca. Esta discrepancia con la bibliografía puede atribuirse a que los resultados de Bahorun *et al.* (2004) se refieren al pimiento picante y no a una variedad dulce¹⁵².

Los valores de capacidad antioxidante obtenidos para la cebolla roja estuvieron entre 4.0 a 5.6 mmol eq. Trolox/100 g (TEAC) y de 5.1 a 6.2 mmol eq. FeSO₄/100 g (FRAP). Estos datos son similares a los obtenidos en el estudio de Gorinstein *et al.* (2009), en el que la cebolla roja tuvo la capacidad antioxidante más alta. Sin embargo, Li *et al.* (2012) obtuvieron que la capacidad antioxidante de la cebolla roja fue menor que la obtenida por otros vegetales caracterizados también por su alta pigmentación (roja o violeta)^{153,154}.

Con respecto a los resultados de ORAC, las muestras que mostraron la capacidad antioxidante más alta fueron las variedades de ajo rojo y blanco, la cebolla roja y el espárrago verde, en este orden (Figura 73B). Mientras que las variedades de ajo tuvieron altos valores ORAC, la capacidad antioxidante de estas muestras conforme a los ensayos TEAC y FRAP fueron bajos (de 0.13 a 0.59 mmol eq. Trolox/100 g y de 0.39 a 0.83 mmol eq. FeSO₄/100 g, respectivamente). Según los resultados de ORAC, no se puede afirmar que la capacidad antioxidante del ajo blanco sea claramente más alta que la variedad de ajo rojo porque el rango mostrado por el ajo blanco es más amplio que el rango de actividad obtenido para el ajo rojo. Este hecho revela que la capacidad antioxidante del ajo blanco varía mucho entre las muestras recolectadas a

¹⁵² Bahorun, T. *et al.*, 2004. Total phenol, flavonoid, proanthocyanidin and vitamin C levels and antioxidant activities of Mauritian vegetables. *Journal of the Science of Food and Agriculture*, 84(12), pp.1553–1561.

¹⁵³ Gorinstein, S. *et al.*, 2009. A comparative study of phenolic compounds and antioxidant and antiproliferative activities in frequently consumed raw vegetables. *European Food Research and Technology*, 228(6), pp.903–911.

¹⁵⁴ Li, H. *et al.*, 2012. Highly pigmented vegetables: Anthocyanin compositions and their role in antioxidant activities. *Food Research International*, 46(1), pp.250–259.



diferentes tiempos en el mismo año. El espárrago verde mostró valores TEAC en un rango de 3.02 a 3.9 mmol eq. Trolox/100 g. Estos datos concuerdan con los resultados de Tiveron *et al.* (2012), donde se estudió la actividad antioxidante de especies vegetales brasileñas, obteniendo que el espárrago mostró un valor de actividad antioxidante considerablemente alto¹⁵⁵. Con respecto a los diferentes tiempos de recolección del espárrago, se detectó muy poca variación entre las muestras (Tabla 2). Se ha de destacar que la fecha de recolección temprana de Febrero correspondió con el valor de capacidad antioxidante más bajo en los ensayos TEAC y FRAP¹⁴⁸.

La capacidad antioxidante del haba puede también destacarse, siendo la quinta y sexta muestra con el resultado más alto en TEAC (de 4.05 a 5.4 mmol eq. Trolox/100 g) y en ORAC (de 1.29 a 2.53 mmol eq. Trolox/100 g), respectivamente. En lo referente a la estacionalidad y el consumo de habas, la fecha más adecuada es Enero. Este hecho puede explicar que el valor más alto de TEAC correspondió con la muestra recolectada ese mes¹⁴⁸.

El vegetal que mostró la capacidad antioxidante más baja fue la patata "Vivaldi" en todos los métodos analizados. Los valores ORAC para esta variedad fueron de 0.238 a 0.286 mmol eq. Trolox/100 g, menor capacidad antioxidante que la obtenida por la patata "Scarlet" con el mismo método (0.550 a 0.75 mmol eq. Trolox/100 g). En base a estudios previos, era de esperar que la capacidad antioxidante fuera más baja para "Vivaldi" que para "Scarlet roja". Hu *et al.* (2012) sugerían que las variedades de patata roja o violeta tienen una mayor capacidad antioxidante que aquellas variedades blancas o amarillas, indicando que los antocianos son importantes contribuyentes a dicha propiedad bioactiva

¹⁵⁵ Tiveron, A.P. *et al.*, 2012. Antioxidant activity of brazilian vegetables and its relation with phenolic composition. *International Journal of Molecular Sciences*, 13(7), pp.8943–8957.



¹⁵⁶. A pesar de que el tercer resultado más alto en los ensayos FRAP se obtuvo para la berenjena larga (5.3-8.7 mmol eq. FeSO₄/100 g), esta misma variedad de berenjena mostró el segundo valor más bajo de capacidad antioxidante en el ensayo ORAC (0.38-0.59 mmol eq. Trolox/100 g). Estos valores obtenidos para la berenjena fueron similares a los proporcionados por Stangeland *et al.* (2009), situándose dicha muestra justo por debajo de la patata dulce en cuanto a su capacidad antioxidante obtenida por el método ORAC¹⁵⁷. La berenjena está disponible todo el año, excepto en verano. Esto puede explicar los bajos valores de capacidad antioxidante obtenidos con los métodos FRAP y ORAC cuando la recolección fue en Septiembre¹⁴⁸.

A partir de los resultados de capacidad antioxidante obtenidos en este estudio para frutas y vegetales, se seleccionaron las muestras de pimiento para realizar un estudio exhaustivo de los componentes polifenólicos.

1.2. Caracterización de polifenoles y otros compuestos polares en tres muestras de pimiento (*Capsicum annun L.*).

El pimiento (*Capsicum annuum L.*), de la familia de las solanáceas, es un cultivo importante en agricultura con numerosas variedades cultivadas alrededor del mundo, no sólo debido a su importancia económica, sino también por el valor nutricional de sus frutos^{158,159}. Muchos estudios acerca del perfil fenólico de *C. annuum* han descrito el contenido de capsacinoides como capsaicina, dihidrocapsaicina y nor-dihidrocapsaicina, compuestos característicos en el pimiento picante, el cual normalmente se consume sólo en Sudamérica. Sin

¹⁵⁶ Hu, C. *et al.*, 2012. Influence of cultivar and year on phytochemical and antioxidant activity of potato (*Solanum tuberosum L.*) in Ontario. *Canadian Journal of Plant Science*, 92, pp.485–493.

¹⁵⁷ Stangeland, T., Remberg, S.F. & Lye, K. a., 2009. Total antioxidant activity in 35 Ugandan fruits and vegetables. *Food Chemistry*, 113(1), pp.85–91.

¹⁵⁸ Materska, M. *et al.*, 2003. Isolation and structure elucidation of flavonoid and phenolic acid glycosides from pericarp of hot pepper fruit *Capsicum annuum L.* *Phytochemistry*, 63(8), pp.893–898.

¹⁵⁹ Navarro, J. *et al.*, 2006. Changes in the contents of antioxidant compounds in pepper fruits at different ripening stages, as affected by salinity. *Food Chemistry*, 96(1), pp.66–73.



embargo, hay pocos estudios que incluyan el estudio de otros metabolitos importantes en este vegetal, especialmente cuando se trata de variedades dulces de pimiento^{160,161}. Por ello, el principal objetivo de este trabajo fue llevar a cabo una caracterización completa de compuestos bioactivos polares presentes en extractos de pimiento de tres variedades diferentes, analizadas mediante el uso de cromatografía líquida de alta resolución en fase inversa (RP-HPLC) acoplada a otros sistemas de detección como el diodo array (DAD) y el analizador de masas de tiempo de vuelo (TOF-MS).

En cuanto a los resultados obtenidos en este estudio, en primer lugar se ha de destacar que la extracción con metanol/agua, como sistema de extracción no selectivo, constituyó un protocolo reproducible. Esta mezcla de disolventes fue capaz de extraer con éxito metabolitos polares y semi-polares de la muestra. Además, al analizar el fruto entero, se encuentran muchos compuestos de diferente polaridad¹⁵⁸. En segundo lugar, la combinación de la eficiencia de la separación por cromatografía líquida con la espectrometría de masas permitió identificar una gran cantidad de metabolitos. El análisis HPLC-DAD-ESI-TOF-MS tuvo como resultado una adecuada separación de las diferentes clases de metabolitos, incluidas algunas formas isoméricas. El tiempo de retención de cada pico, como un indicativo de polaridad, permitió inferir el tipo de metabolito de acuerdo a la región del cromatograma en la cual eluye cada familia. Además, la detección UV-vis dada por el detector de batería de diodos (DAD) permitió la diferenciación entre las familias de estos compuestos dado que cada familia tiene bandas de absorción características.

¹⁶⁰ Serrano, M. *et al.*, 2010. Antioxidant and nutritive constituents during sweet pepper development and ripening are enhanced by nitrophenolate treatments. *Food Chemistry*, 118(3), pp.497–503.

¹⁶¹ Ornelas-Paz, J.D.J. *et al.*, 2010. Effect of cooking on the capsaicinoids and phenolics contents of Mexican peppers. *Food Chemistry*, 119(4), pp.1619–1625.



El método TOF-MS proporcionó alta resolución, medida de masa exacta e información de la composición elemental. Complementariamente, algunos compuestos fueron fragmentados en la fuente del analizador TOF, proporcionando el patrón de fragmentación por la disociación en la fuente inducida por colisión (ISCID) en el espectro de masas.

Los perfiles de los tres extractos de pimiento analizados se muestran en la Figura 74, donde la intensidad de la señal aparece con la misma escala en todos los cromatogramas. Los tres perfiles muestran los cromatogramas de pico base (BPC) obtenidos mediante HPLC-ESI-TOF-MS para las tres variedades analizadas (A) pimiento verde, (B) pimiento amarillo y (C) pimiento rojo. Como puede verse tanto en la Figura 74 como en la Tabla 19 y la Tabla 20, los compuestos identificados están numerados del 1 al 45, mientras que los compuestos desconocidos aparecen nombrados en letra mayúscula de la A a la U.

Cabe señalar que en el perfil (A), correspondiente a la variedad ecológica, los picos son de mayor intensidad que en las otras variedades de pimiento (B) y (C).

Se identificó de manera tentativa un total de 45 compuestos, pertenecientes a varias familias que incluían ácidos orgánicos, aminoácidos, nucleósidos, ácidos fenólicos (hidroxibenzoico, hidroxicinámico y ácidos fenilacéticos), alcoholes fenólicos, flavonoides y lignanos. En la tabla se muestran todos los compuestos propuestos, agrupados por categoría de metabolito y detallando el tiempo de retención, la banda UV (nm), la masa exacta m/z $[M-H]^-$ experimental y calculada, el patrón de fragmentación ISCID, el error y los valores $mSigma$, así como la fórmula molecular y la identificación propuesta en base a la fórmula. La Tabla 20 muestra una lista de compuestos para los cuales no fue posible dilucidar una estructura porque no hubo suficientes evidencias para llegar a una propuesta de estructura. Una completa explicación de los compuestos



identificados tentativamente se presenta a continuación en el mismo orden en mostrado en la Tabla 19.

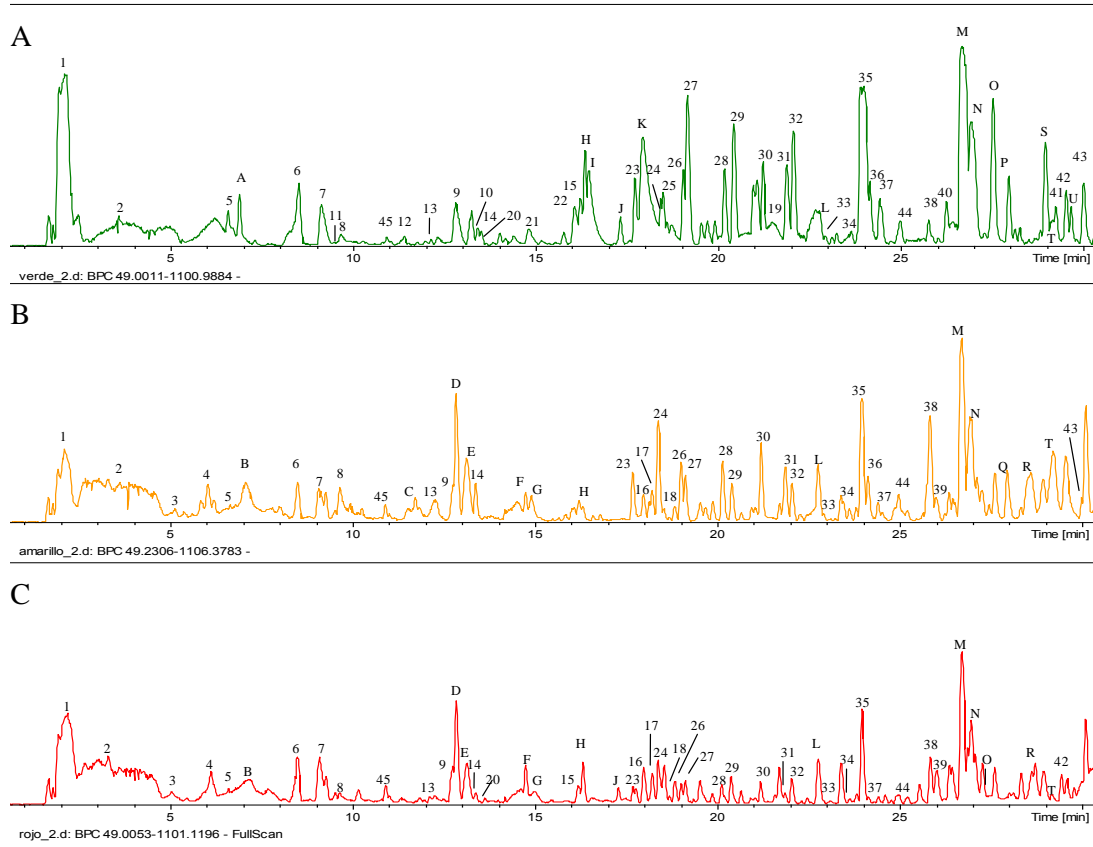


Figura 74. Cromatogramas de pico base de los tres extractos de pimienta analizados mediante HPLC-ESI-TOF-MS: A (pimiento verde), B (pimiento amarillo) y C (pimiento rojo).

Tabla 19. Compuestos identificados en los extractos de pimiento analizados mediante HPLC-ESI-TOF-MS: A (pimiento verde), B (pimiento amarillo) y C (pimiento rojo).

#	LC RT (min)	DAD UV (nm)	ESI (-) TOF			ISCID	Error (ppm)	mσ	Fórmula Molecular	Identificación tentativa Compuestos	Extractos		
			Exp. [M-H] ⁻ (Da)	Calc. [M-H] ⁻ (Da)							A	B	C
<i>Ácidos orgánicos</i>													
1	2,23	221	191,0581	191,0561	-	-10,3	14,7	C ₇ H ₁₁ O ₆	Ácido quínico	+	+	+	
2	3,33	243	191,0206	191,0197	-	-4,5	4,2	C ₆ H ₇ O ₇	Ácido cítrico	+	+	+	
3	5,05	219, 304	175,0262	175,0248	-	-8,2	0,5	C ₆ H ₇ O ₆	Ácido ascórbico (isómero I)	-	+	+	
4	6,14	224, 281	175,0258	175,0248	-	-5,9	4,6	C ₆ H ₇ O ₆	Ácido ascórbico (isómero II)	-	+	+	
<i>Nucleósidos</i>													
5	6,61	221, 262	243,0634	243,0623	200,0599 [M-H-CHNO] ⁻	-4,5	16,5	C ₉ H ₁₁ N ₂ O ₆	Uridina	+	+	+	
6	8,48	220, 257	282,0852	282,0844	150,0444 [M-H-pent] ⁻	-3	5,6	C ₁₀ H ₁₂ N ₅ O ₅	Guanosina	+	+	+	
<i>Amino ácidos y derivados</i>													
7	9,1	217	164,0721	164,0717	-	-2,5	4,4	C ₉ H ₁₀ NO ₂	DL-fenilalanina	+	+	+	
8	9,65	219, 295	326,1249	326,1245	164,0732 [M-H-hex] ⁻	-1	11,5	C ₁₅ H ₂₀ NO ₇	Fructosil fenilalanina	+	+	+	
9	12,73	223, 278	203,0834	203,0826	-	-3,8	5,6	C ₁₁ H ₁₁ N ₂ O ₂	Triptófano	+	+	+	
<i>Derivados de fenoles simples</i>													
10	13,4	224, 296	403,1244	403,1246	223,0605[M-H-hex-H ₂ O] ⁻	0,6	23,2	C ₁₇ H ₂₃ O ₁₁	2'-O-β-Apiofuranosil arbutina	+	-	-	
<i>Ácidos hidroxibenzoicos y derivados</i>													
11	9,47	217	299,0751	299,0772	-	7,1	15,3	C ₁₃ H ₁₅ O ₈	Ácido p-hidroxibenzoico β-glucósido (isómero I)	+	-	-	
12	11,39	219	329,0871	329,0878	167,0369 [M-H-hex] ⁻	2	20,4	C ₁₄ H ₁₇ O ₉	Ácido vanílico 1-O-β-D-glucopiranosilester (isómero I)	+	-	-	
13	12,13	220, 258	299,0765	299,0772	-	2,6	22,8	C ₁₃ H ₁₅ O ₈	Ácido p-hidroxibenzoico β-glucósido (isómero II)	+	+	+	



14	13,38	224	329,0872	329,0878	-	1,9	7,4	C ₁₄ H ₁₇ O ₉	Ácido vanílico 1-O-β-D-glucopiranosilester (isómero II)	+	+	+
<i>Ácidos hidrocínámicos y derivados</i>												
15	16,21	223, 280	341,0883	341,0878	179,0378 [M-H-hex] ⁻	-1,5	2,1	C ₁₅ H ₁₇ O ₉	Ácido caféico 4-O-β-D-glucopiranosido o 1-O-Cafeoil-β-D-glucopiranosido (isómero I)	+	-	+
16	17,95	240, 329	355,1051	355,1035	175,0408 [M-H-hex-H ₂ O] ⁻	-4,7	24,9	C ₁₆ H ₁₉ O ₉	Feruloil glucosa (isómero I)	-	+	+
17	18,17	240, 329	385,1145	385,114	205,0517 [M-H-hex-H ₂ O] ⁻	-1,2	3,3	C ₁₇ H ₂₁ O ₁₀	Ácido sinápico-O-hexósido	-	+	+
18	18,65	234	355,1037	355,1035	175,0354 [M-H-hex-H ₂ O] ⁻	-0,8	15,9	C ₁₆ H ₁₉ O ₉	Feruloil glucosa (isómero II)	-	+	+
19	21,43	232, 270, 333	341,0913	341,0878	-	-10,4	25,7	C ₁₅ H ₁₇ O ₉	Ácido caféico 4-O-β-D-glucopiranosido o 1-O-Cafeoil-β-D-glucopiranosido (isómero II)	+	-	-
<i>Ácidos fenilacéticos y derivados</i>												
20	13,47	223	343,1043	343,1035	-	-2,5	22,1	C ₁₅ H ₁₉ O ₉	Ácido homovanílico hexosa	+	-	+
<i>Flavonoides y glicósidos</i>												
21	14,81	233, 284	627,1576	627,1567	-	-2	2,7	C ₂₇ H ₃₁ O ₁₇	7-O-β-D-Glucopiranosil-dihidroquercetina-3-O-α-D-glucopiranosido	+	-	-
22	16,01	232, 289	597,1474	597,1461	-	-3,7	7,08	C ₂₆ H ₂₉ O ₁₆	Quercetina 3-vicianósido	+	-	-
23	17,68	260, 345	609,1488	609,1461	-	-4,4	13,4	C ₂₇ H ₂₉ O ₁₆	Quercetina dihexósido (isómero I)	+	+	+
24	18,33	254, 350	609,1504	609,1461	445,2090 [M-H-hex-2H] ⁻	-7,1	32,6	C ₂₇ H ₂₉ O ₁₆	Quercetina dihexósido (isómero II)	+	+	+
25	18,47	241	741,1919	741,1884	609,1482 [M-H-pent] ⁻	1,8	14,3	C ₃₂ H ₃₇ O ₂₀	Rutina pentósido	+	-	-
26	18,99	269, 340	579,1382	579,1355	-	-4,7	22,4	C ₂₆ H ₂₇ O ₁₅	Luteolina 6-C-β-D-glucopiranosido-8-C-α-L-arabinopiranosido o Luteolina 7-O-[2-(β-d-apiofuranosil)-β-D- arabinopiranosido] (isómero I)	+	+	+
27	19,09	232, 267	593,1582	593,1512	-	-11,8	31,7	C ₂₇ H ₂₉ O ₁₅	Quercetina 3,7-di-O-α-L-rhamnopiranosido	+	+	+
28	20,14	272, 345	579,1402	579,1355	-	-8	5	C ₂₆ H ₂₇ O ₁₅	Luteolina 6-C-β-D-glucopiranosido-8-C-α-L-arabinopiranosido o Luteolina 7-O-[2-(β-d-apiofuranosil)-β-D- arabinopiranosido] (isómero II)	+	+	+
29	20,37	267, 330	563,1459	563,1406	-	-9,4	38,5	C ₂₆ H ₂₇ O ₁₄	Apigenina 7-O-β-D-apiofuranosil(1→2)-β-D-glucopiranosido o Apigenina 8-C-α-L-arabinósido 6-C-β-D-glucósido (isómero I)	+	+	+
30	21,16	269, 344	447,0961	447,0933	-	-6,3	1,5	C ₂₁ H ₁₉ O ₁₁	Quercetina 3-rhamnopiranosido o Luteolina-8-glucósido (isómero I)	+	+	+
31	21,83	265, 345	447,0965	447,0933	-	-7,3	13,3	C ₂₁ H ₁₉ O ₁₁	Quercetina 3-rhamnopiranosido o Luteolina-8-glucósido (isómero II)	+	+	+
32	22,01	270, 330	563,1467	563,1406	-	-10,8	23,5	C ₂₆ H ₂₇ O ₁₄	Apigenina 7-O-β-D-apiofuranosil(1→2)-β-D-glucopiranosido o Apigenina 8-C-α-L-arabinósido 6-C-β-D-glucósido (isómero II)	+	+	+



33	22,88	228, 262	433,1135	433,114	-	1,3	20,4	C ₂₁ H ₂₁ O ₁₀	Naringenina chalcona hexosa (isómero I)	+	+	+
34	23,59	247, 343	433,1135	433,114	217,0353 [M-H-hex] ⁻	1,3	29,5	C ₂₁ H ₂₁ O ₁₀	Naringenina chalcona hexosa (isómero II)	+	+	+
35	23,82	252, 347	579,1431	579,1508	-	13,4	32,8	C ₃₀ H ₂₇ O ₁₂	Naringenina-7-O-β-D-(3''-p-coumaroil)- glucopiranosido	+	+	+
36	24,1	238, 323	597,1848	597,1827	273,1333 [M-H-2hex] ⁻	-3,9	9,5	C ₂₇ H ₃₃ O ₁₅	Phloretin dihexósido	+	+	+
37	24,35	250, 346	463,0899	463,0882	-	-3,7	3,5	C ₂₁ H ₁₉ O ₁₂	Quercetina 3'-O-β-D-glucopiranosido o Quercetina 4'-β-D- glucopiranosido	+	+	+
38	25,76	320	609,1494	609,1461	-	-5,4	43,5	C ₂₇ H ₂₉ O ₁₆	Quercetina dihexósido (isómero III)	+	+	+
39	25,96	250	827,1935	827,1888	621,1440 [M-H-hex-CO ₂] ⁻	-5,7	14,8	C ₃₅ H ₃₉ O ₂₃	Luteolina 7-O-(2-apiofuranosil-4-glucopiranosil-6-malonil) glucopiranosido	-	+	+
40	26,18	232, 280	565,1573	565,1563	-	-1,8	5,6	C ₂₆ H ₂₉ O ₁₄	Teaflavanosido I	+	-	-
41	29,16	266	431,0993	431,0983	-	-2,2	1,2	C ₂₁ H ₁₉ O ₁₀	Apígenina 8-C-glucósido	+	-	-
42	29,44	NA	461,1094	461,1089	161,0470 [M-H-hex] ⁻	-1,1	7	C ₂₂ H ₂₁ O ₁₁	Diosmetina 7-O-β-D- glucósido	+	-	+
43	29,93	239, 366	433,1164	433,114	271,0630 [M-H-hex] ⁻	-5,4	29,1	C ₂₁ H ₂₁ O ₁₀	Naringenina 7-O-β-D- glucósido	+	+	-
<i>Lignanos y derivados</i>												
44	24,92	NA	521,2022	521,2028	-	1,3	4,8	C ₂₆ H ₃₃ O ₁₁	Lariciresinol glucopiranosido	+	+	+
<i>Otros metabolitos</i>												
45	10,92	219, 295	218,105	218,1034	-	-7,2	5,3	C ₉ H ₁₆ NO ₅	Ácido pantoténico	+	+	+

^a "NA" Absorbancia no detectada o menor que 5 mAU.

^b "Exp. y Calc." masa exacta experimental y calculada.

^c "mσ" valor mSigma.

^d "-" Fragmentación en la fuente no detectada.

^e Hex, pérdida de hexosa (-162 u); pent, pérdida de pentosa (-132 u).

^f Determinación cualitativa en cada extracto de pimienta: (+) para compuestos detectados (-) para compuestos no detectados.



Tabla 20. Compuestos desconocidos en los tres extractos de pimienta.

#	LC RT (min)	ESI (-)		Error (ppm)	$m\sigma$	Fórmula molecular	Extractos de pimienta		
		Exp. [M-H] ⁻ (Da)	Calc. [M-H] ⁻ (Da)				Italiano verde ecológico	"Lamuyo" amarillo	"California" rojo
A	6,91	611,1479	611,1465	-2,3	38,80	C ₂₃ H ₃₁ O ₁₉	+	-	-
B	7,06	241,0359	241,0354	-2,1	1,50	C ₁₀ H ₉ O ₇	-	+	+
C	11,71	281,0767	281,0819	18,8	12,10	C ₁₇ H ₁₃ O ₄	-	+	-
D	12,83	223,0264	223,0248	-7,3	10,90	C ₁₀ H ₇ O ₆	-	+	+
E	13,12	381,0475	381,0463	-3,1	12,00	C ₁₆ H ₁₃ O ₁₁	-	+	+
F	14,72	443,1935	443,1923	-2,8	34,50	C ₂₁ H ₃₁ O ₁₀	-	+	+
G	14,89	267,0158	267,0146	-4,4	13,30	C ₁₁ H ₇ O ₈	-	+	+
H	16,33	651,1988	651,1931	-8,8	15,00	C ₃₀ H ₃₅ O ₁₆	+	+	+
I	16,43	611,1663	611,1618	-7,5	24,50	C ₂₇ H ₃₁ O ₁₆	+	-	-
J	17,26	406,1681	406,1660	-2,1	38,80	C ₂₄ H ₂₄ NO ₅	+	-	+
K	17,90	581,1560	581,1560	-8,3	32,30	C ₂₆ H ₂₉ O ₁₅	+	-	-
L	22,41	359,1480	359,1480	5,5	4,60	C ₂₀ H ₂₃ O ₆	+	+	+
M	26,60	447,1024	447,1085	19,4	15,80	C ₂₅ H ₁₉ O ₈	+	+	+
N	26,88	621,1528	621,1528	13,8	10,80	C ₃₂ H ₂₉ O ₁₃	+	+	+
O	27,45	187,1019	187,0976	-3,9	17,40	C ₉ H ₁₅ O ₄	+	-	+
P	27,88	243,1257	243,1238	-7,8	4,30	C ₁₂ H ₁₉ O ₅	+	-	-
Q	27,87	675,3255	675,3255	-3,2	5,80	C ₃₂ H ₅₁ O ₁₅	-	+	-
R	28,52	923,3781	923,3781	-1,6	5,60	C ₄₁ H ₆₃ O ₂₃	-	+	+
S	28,89	539,1817	539,1817	-8,6	11,80	C ₂₅ H ₃₁ O ₁₃	+	-	-
T	29,12	761,3265	761,3265	-3,7	18,30	C ₃₅ H ₅₃ O ₁₈	+	+	+
U	29,59	545,2632	545,2632	-5,2	14,60	C ₂₆ H ₄₁ O ₁₂	+	-	-

^a"Exp. y Calc." masa exacta experimental y calculada.

^b" $m\sigma$ " valor *mSigma*.

^cDeterminación cualitativa en cada extracto de pimienta: (+) para compuestos detectados (-) para compuestos no detectados.

El primer compuesto en el BPC fue tentativamente identificado como ácido quínico, correspondiente al pico 1 en el tiempo de retención 2.23 min y m/z 191.0581. Este compuesto, por lo que se conoce en bibliografía, no había sido identificado en *C. annuum* L. previamente. Mostró absorción UV a 221 nm de longitud de onda, lo que concuerda con las bandas típicas de UV de los ácidos



orgánicos de otras matrices vegetales encontradas en bibliografía. El pico 2, en el tiempo de retención de 3.33 min y m/z 191.0206 se pudo asignar al ácido cítrico, otro ácido orgánico que no se había descrito previamente descrito en pimiento. Este compuesto se ha encontrado en muchas frutas y vegetales, así como en el gazpacho, que contiene tomate y pimiento, dos vegetales de la familia *Solanaceae*^{162,163}.

Dos formas isoméricas del ácido ascórbico (picos 3 y 4) eluyeron a 5.05 min y 6.4 min con m/z de 175.0262 y 175.0258, respectivamente. La presencia de estos compuestos en pimiento ha sido descrita por Jung *et al.* (2010). Es importante indicar que el extracto de pimiento verde es el único de los extractos analizados de pimiento que no contiene ácido ascórbico, más comúnmente conocido como vitamina C¹⁶⁴.

El pico 5 con m/z 243.0634 y tiempo de retención 6.61 min, mostró un fragmento a 200 m/z ($[M-H-CHNO]^-$). Este compuesto se identificó como uridina¹⁶⁵. El siguiente compuesto identificado como adenosina (pico 6) se encontró a 8.48 min y m/z 282.0852 y su patrón de fragmentación presentó un fragmento a 150 m/z, correspondiente a la pérdida de una fracción de pentosa. Ésta es la primera vez que estos dos nucleósidos son descritos en pimiento, habiéndose encontrado en todas las muestras analizadas. La asignación de estos

¹⁶² Flores, P., Hellín, P. & Fenoll, J., 2012. Determination of organic acids in fruits and vegetables by liquid chromatography with tandem-mass spectrometry. *Food Chemistry*, 132(2), pp.1049–1054.

¹⁶³ Verde Mendez, C.M., *et al.*, 2011. Vitamin C and organic acid contents in Spanish Gazpacho soup related to the vegetables used for its elaboration process. *Cyta-Journal of Food*, 9(1), pp. 71–76.

¹⁶⁴ Jung, M., *et al.*, 2010. Analyses of capsaicinoids and ascorbic acid in pepper (*Capsicum annum* L.) breeding lines. *Journal of the Korean Society of Food Science and Nutrition*, 39(11), pp. 1705–1709.

¹⁶⁵ De Marino, S. *et al.*, 2006. New constituents of sweet *Capsicum annum* L. fruits and evaluation of their biological activity. *Journal of Agricultural and Food Chemistry*, 54(20), pp.7508–7516.



compuestos se hizo posible dada su previa identificación en extractos polifenólicos de tomate¹⁶⁶.

Se encontraron tres aminoácidos y sus derivados en todas las muestras analizadas. El pico 7 en el minuto 9.10 con m/z 164.0721 fue identificado como DL-fenilalanina y el pico 9 en el minuto 12.73 con m/z 203.0834 fue identificado como triptófano. La presencia de estos compuestos en *C. annuum* L. había sido descrita por Ritota *et al.* (2010), específicamente en el pimiento "dulce italiano"¹⁶⁷. El pico 8 con m/z 326.1249 fue detectado al tiempo de retención de 9.65 min. Éste presentó un fragmento de m/z 164, que correspondió con la pérdida de la fracción de hexosa. De acuerdo con esto, el patrón de fragmentación en la fuente y la detección previa de este compuesto en tomate, hizo posible su identificación fructosil fenilalanina¹⁶⁶.

El pico 10 correspondió a un derivado de fenol simple con m/z 403.1244. Este compuesto fue propuesto como 2'-O-β-apiofuranosil arbutina tras la comparación con la fórmula molecular C₁₇H₂₃O₁₁ ([M-H]⁻) en las bases de datos y teniendo en cuenta el fragmento mostrado a m/z 223 debido a la pérdida de un azúcar¹⁶⁸. Esta es la primera vez que este fenol es identificado en pimiento, aunque fue detectado sólo en la variedad de pimiento verde.

Con respecto al ácido hidroxibenzoico y sus derivados, dos formas isoméricas del ácido p-hidroxibenzoico β-glucósido correspondieron a los picos 11 y 13, los cuales eluyeron a los tiempos de 9.47 min y 12.13 min y presentaron una masa exacta ([M-H]⁻) de 299.0751 y 299.0765, respectivamente. De manera similar,

¹⁶⁶ Gómez-Romero, M., Segura-Carretero, A. & Fernández-Gutiérrez, A., 2010. Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. *Phytochemistry*, 71(16), pp.1848–1864.

¹⁶⁷ Ritota, M. *et al.*, 2010. Metabolomic characterization of Italian sweet pepper (*Capsicum annuum* L.) by means of HRMAS-NMR spectroscopy and multivariate analysis. *Journal of Agricultural and Food Chemistry*, 58(17), pp.9675–9684.

¹⁶⁸ Ahmed, A.S., 2006. Phytochemical and biological study of *Grevillea robusta* A. Cunn cultivated in Egypt. *Bulletin of Pharmaceutical Sciences*, 29, pp. 272–288.



se identificaron dos isómeros del ácido vanílico 1-O-β-D-glucopiranosilester, correspondiendo con los picos 12 y 14. El primer isómero eluído en el minuto 11.39 presentó una masa de m/z 329.0871 y el segundo isómero eluído a los 13.38 min difirió en sólo 0.1 mDa. El pico 12 también generó un fragmento de m/z 167, correspondiente a la pérdida del grupo azúcar. Es la primera vez que estos cuatro ácidos hidroxibenzóicos se describen en muestras de pimienta.

La identificación de los metabolitos se basó en su presencia en otros vegetales como *Solanum lycopersicum* L. y *Cydonia oblonga* L. en el caso del ácido p-hidroxibenzóico β-glucósido, y *Cucumis sativus* L. en el caso del ácido vanílico 1-O-β-D-glucopiranosiléster^{166,169,170}. Todos estos compuestos fueron encontrados en el extracto de pimienta "dulce italiano", mientras que en los otros extractos sólo se encontraron las últimas formas isoméricas de cada compuesto.

Con respecto al ácido hidroxicinámico y derivados, se encontraron cinco compuestos. Los picos 15 y 19 fueron tentativamente identificados como ácido caféico 4-O-β-D-glucopiranosido o 1-O-cafeoil-β-D-glucopiranosido (ambos son isómeros estructurales). El pico 15 eluyó al minuto 16.2 y presentó una m/z de 341.0883 con un fragmento de m/z 179, patrón de fragmentación característico de la pérdida de una unidad de glucosa. El pico 19 eluyó a 21.43 min con m/z 341.0913, aunque para este segundo isómero no se encontraron fragmentos en la fuente de ionización del espectrómetro de masas. La presencia de estos compuestos no se ha mostrado previamente en *C. annuum* L. pero se han

¹⁶⁹ Fiorentino, A. *et al.*, 2008. Isolation and structure elucidation of antioxidant polyphenols from quince (*Cydonia vulgaris*) peels. *Journal of Agricultural and Food Chemistry*, 56(8), pp.2660–2667.

¹⁷⁰ Abu-Reidah, I.M. *et al.*, 2012. HPLC–ESI-Q-TOF-MS for a comprehensive characterization of bioactive phenolic compounds in cucumber whole fruit extract. *Food Research International*, 46(1), pp.108–117.



encontrado en tomate, vegetal que pertenece a la misma familia del pimiento y también en pepino^{170,171}.

Con relación a la comparación de la presencia de estos compuestos en las muestras analizadas, es importante destacar que ambas formas isoméricas fueron encontradas en el extracto de pimiento verde, mientras que ninguno de los dos isómeros se identificó en el pimiento amarillo y sólo el primer isómero se encontró en la variedad de pimiento rojo. Un caso similar fue el del compuesto identificado como feruloil glucosa, el cual mostró dos formas isoméricas a 17.95 min con m/z 355.1051 (pico 16) y a 18.65 min con m/z 355.1037 (pico 18). Se detectó el fragmento correspondiente a la aglicona (m/z 175) para ambos picos. Park *et al.* (2012) describió la presencia de feruloil glucosa en pimiento amarillo, lo cual coincide con el hecho de haber sido encontrado sólo en variedades de pimiento amarillo y rojo¹⁷². Otro compuesto nuevo se identificó en pimiento en el tiempo de retención 18.17 min y m/z 385.1145 (pico 17), el cual fue disociado a m/z 205 debido a la pérdida de una molécula de azúcar de seis carbonos. Este compuesto se identificó como ácido sinápico-O-hexósido, no siendo detectado en el pimiento "dulce italiano"¹⁶⁶.

Se identificó un compuesto perteneciente al grupo de los fenilacéticos en el minuto 13.46 (pico 20). Éste fue identificado como ácido homovanílico hexosa, siendo la primera vez que este compuesto se describe en pimiento¹⁶⁶. Específicamente, este compuesto se encontró en el extracto analizado de pimiento rojo y también en el pimiento verde, pero no en el extracto de pimiento amarillo.

¹⁷¹ Moco, S. *et al.*, 2006. A liquid chromatography-mass spectrometry-based metabolome database for tomato 1. *Plant Physiology*, 141, pp.1205–1218.

¹⁷² Park, S. *et al.*, 2012. Determination of polyphenol levels variation in *Capsicum annuum* L. cv. Chelsea (yellow bell pepper) infected by anthracnose (*Colletotrichum gloeosporioides*) using liquid chromatography-tandem mass spectrometry. *Food Chemistry*, 130(4), pp.981–985.



Los flavonoides y sus derivados glicósidos constituyen el grupo más importante de compuestos fenólicos encontrados en las muestras de pimiento. Se logró identificar un total de 23 flavonoides, como se expone a continuación. El pico 26 se identificó como luteolina 6-C-β-D-glucopiranosido-8-C-α-L-arabinopiranosido o luteolina 7-O-[2-(β-D-apiofuranosil)-β-D-glucopiranosido] y también se encontró un isómero en el minuto 21.14 (pico 28). También se han propuesto dos posibilidades para las dos formas isoméricas correspondientes a los picos 30 y 31. Comparando su fórmula molecular (C₂₁H₁₉O₁₁), estos compuestos fueron identificados como quercetina 3-ramnopiranosido o luteolina-8-glucósido.

El pico 39 con m/z 827.1935 fue identificado como luteolina 7-O-(2-apiofuranosil-glucopiranosil-6-malonil) glucopiranosido. El fragmento mostrado en el espectro de masas a m/z 621 pudo deberse a la pérdida de hexosa y CO₂. Estos compuestos han sido previamente descritos en *C. annuum* y, en el presente estudio, se han identificado todos ellos en las tres muestras analizadas de pimiento excepto por el pico 39, ausente en la variedad de pimiento verde^{158,165,172-177}. Este hecho concuerda con la bibliografía, habiéndose descrito el compuesto 39 en frutos de pimiento^{158,174} y los compuestos 26 y 30, en variedades de pimiento dulce como el pimiento amarillo cv. Chelsea¹⁷² y la variedad de pimiento ruso Anastasia Black¹⁷⁶.

¹⁷³ Marín, A. *et al.*, 2004. Characterization and quantitation of antioxidant constituents of sweet pepper (*Capsicum annuum* L.). *Journal of Agricultural and Food Chemistry*, 52(12), pp.3861–3869.

¹⁷⁴ Materska, M. & Perucka, I., 2005. Antioxidant activity of the main phenolic compounds isolated from hot pepper fruit (*Capsicum annuum* L.). *Journal of Agricultural and Food Chemistry*, 53(5), pp.1750–1756.

¹⁷⁵ Ochi, T., *et al.*, 2005. Chemical constituents of *Capsicum annuum* L. var. angulosum, and anti Helicobacter pylori activity. *Natural Medicines*, 59(2), pp. 76–84.

¹⁷⁶ Schelz, Z. *et al.*, 2006. Qualitative analysis of MDR-reversing Anastasia Black (Russian black sweet pepper, *Capsicum annuum*, Solanaceae) extracts and fractions by HPLC and LC-MS-MS methods. *In vivo*, 20(5), pp.651–656.

¹⁷⁷ Yanagida, A. *et al.*, 2007. Comprehensive separation of secondary metabolites in natural products by high-speed counter-current chromatography using a three-phase solvent system. *Journal of Chromatography A*, 1151(1-2), pp.74–81.



Todas las muestras analizadas contenían los compuestos 23, 24, 33, 34, 36, 38 y 43, los cuales fueron identificados como siete flavonoides muy conocidos previamente citados por muchos autores en frutas, vegetales o plantas pertenecientes a la familia *Solanaceae*^{166,171,178-180}.

Los picos 23, 24 y 38 como formas isoméricas de m/z 609 en los tiempos de retención 17.68, 18.33 y 25.76 min, respectivamente, fueron identificados como dihexósido de quercetina. Sólo el pico 24 mostró un fragmento de m/z 445, correspondiente a la pérdida de un azúcar. El pico 33 en el minuto 22.88 y el pico 34 a 23.59 min fueron identificados como formas isoméricas de la naringenina chalcona hexosa. Sólo el segundo isómero mostró un fragmento a m/z 217, correspondiente a la pérdida de la hexosa. El pico 36 fue identificado como phloretin dihexósido, que mostró un fragmento de m/z 273, correspondiente a la forma aglicona del compuesto. El pico 43 eluyó a 29.93 min con m/z 433.1164, mostrando un fragmento de m/z 271 (glucosa) y fue identificado como naringenina 7-O- β -D-glucósido.

El resto de compuestos identificados de la familia de los flavonoides (picos 21, 22, 25, 27, 29, 32, 35, 37, 40, 41 y 42), son descritos por primera vez en esta matriz, siendo posible su identificación gracias a su presencia en otras plantas como la espinaca, la cebolla, el té rooibos, etc^{178,179,181,182}. Cinco de estos compuestos en éste se encontraron solo en el extracto de pimiento "dulce

¹⁷⁸ Helmja, K. *et al.*, 2007. Characterization of bioactive compounds contained in vegetables of the Solanaceae family by capillary electrophoresis. *Proceedings of the Estonian Academy of Sciences Chemistry*, 56(4), pp.172–186.

¹⁷⁹ Niranjana, A. *et al.*, 2011. Development and optimization of HPLC-PDA-MS-MS method for simultaneous quantification of three classes of flavonoids in legume seeds, vegetables, fruits, and medicinal plants. *Journal of Liquid Chromatography & Related Technologies*, 34(16), pp.1729–1742.

¹⁸⁰ Vallverdú-Queralt, A. *et al.*, 2011. Screening of the polyphenol content of tomato-based products through accurate-mass spectrometry (HPLC-ESI-QTOF). *Food Chemistry*, 129(3), pp.877–883.

¹⁸¹ Iswaldi, I. *et al.*, 2011. Identification of phenolic compounds in aqueous and ethanolic rooibos extracts (*Aspalathus linearis*) by HPLC-ESI-MS (TOF/IT). *Analytical and Bioanalytical Chemistry*, 400(10), pp.3643–3654.

¹⁸² Prasad, K.N. *et al.*, 2005. Isolation of a free radical-scavenging antioxidant from water spinach (*Ipomoea aquatica* Forsk). *Journal of the Science of Food and Agriculture*, 85(9), pp.1461–1468.



italiano". Los iones con m/z 627.1576 y 597.1474 se identificaron como 7-O- β -D-glucopiranosil-dihidroquercetina-3-O- α -D-glucopiranosido (pico 21 a 14.81 min) y quercetina 3-vicianósido (pico 22 a 16.01 min). Basándonos en los de datos de espectrometría de masas, el pico 25 a 18.47 min con m/z 741.1919 y su fragmento a m/z 609 confirmó la presencia de rutina pentósido. Este hecho concuerda con la información encontrada en la bibliografía¹⁸³. Los compuestos con iones $[M-H]^-$ a m/z 565.1573 (pico 40) y 431.0993 (pico 419) fueron identificados como teaflavánosido I y apigenina 8-C-glucósido, respectivamente^{184,185}. Se detectaron en todas las muestras estudiadas en este trabajo cinco flavonoides. El pico 27 fue tentativamente identificado como quercetina 3-ramnopiranosido¹⁸⁶. Los picos 29 y 32 a 20.37 min y 22.01 min fueron isómeros y se identificaron como apigenina 7-O- β -D-apiofuranosil (1 \rightarrow 2)- β -D-glucopiranosido o apigenina 8-C- α -L-arabinosido 6-C- β -D-glucósido^{187,188}. El pico 35 a m/z 579.1431 fue identificado como narigenina-7-O- β -D-(3"-p-cumaroil)-glucopiranosido¹⁸⁹. Para el ión $[M-H]^-$ a 463.0899 (pico 37) se sugirieron dos posibles estructuras, quercetina 3'-O- β -D-glucopiranosido o quercetina 4'- β -D-glucopiranosido¹⁹⁰. Finalmente, se encontró en todas las muestras excepto en el pimiento "Lamuyo" un nuevo flavonoide, que fue

¹⁸³ Lin, L.-Z., Chen, P. & Harnly, J.M., 2008. New phenolic components and chromatographic profiles of green and fermented teas. *Journal of Agricultural and Food Chemistry*, 56(17), pp.8130–8140.

¹⁸⁴ Chen, J.-H. *et al.*, 2010. Identification and evaluation of antioxidants defatted *Camellia oleifera* seeds by isopropanol salting-out pretreatment. *Food Chemistry*, 121(4), pp.1246–1254.

¹⁸⁵ Sakakibara, H. *et al.*, 2003. Simultaneous determination of all polyphenols in vegetables, fruits, and teas. *Journal of Agricultural and Food Chemistry*, 51(3), pp.571–581.

¹⁸⁶ Sugimoto, T. *et al.*, 2011. Use of ultra-performance liquid chromatography/time-of-flight mass spectrometry with nozzle-skimmer fragmentation for comprehensive quantitative analysis of secondary metabolites in *Arabidopsis thaliana*. *Journal of Separation Science*, 34(24), pp.3587–3596.

¹⁸⁷ Dinelli, G. *et al.*, 2011. Profiles of phenolic compounds in modern and old common wheat varieties determined by liquid chromatography coupled with time-of-flight mass spectrometry. *Journal of Chromatography A*, 1218(42), pp.7670–7681.

¹⁸⁸ Scoparo, C.T. *et al.*, 2012. Analysis of *Camellia sinensis* green and black teas via ultra high performance liquid chromatography assisted by liquid-liquid partition and two-dimensional liquid chromatography (size exclusion \times reversed phase). *Journal of Chromatography A*, 1222, pp.29–37.

¹⁸⁹ Yuan, W., 2007. Flavonoids, coumarins and Triterpenes from the aerial parts of *Cnidocolus texanus*. *Planta Medica*, 73(12), pp. 1304–1308.

¹⁹⁰ Lu, X. *et al.*, 2011. Determination of total phenolic content and antioxidant capacity of onion (*Allium cepa*) and shallot (*Allium oschaninii*) using infrared spectroscopy. *Food Chemistry*, 129(2), pp.637–644.



identificado a 29.44 min y m/z 461.1094 (pico 42). Este compuesto se propuso como diosmetina 7-O-β-D-glucósido debido a la presencia en el espectro de masas a ese tiempo de retención de un fragmento correspondiente a la diosmetina¹⁹¹.

Con respecto a los lignanos y derivados, se identificó un compuesto en todas las muestras de pimienta analizadas. Este compuesto corresponde al pico 44 a 24.92 min y m/z 521.2022, cuya identificación como lariciresinol glucopiranosido ha sido citada previamente por Funari *et al.* (2011)¹⁹². Otro metabolito presente en todas las variedades de pimienta se identificó como ácido pantoténico (pico 45) a 10.92 min y m/z 218.1050. Ésta fue la primera vez que estos dos compuestos han sido identificados en muestras de pimienta¹⁷⁸. Del total de compuestos identificados, 35 son polifenoles, siendo los flavonoides la principal familia de compuestos, sobre todo especies glicosiladas o flavonoles. De los metabolitos identificados, 11 han sido descritos en pimienta anteriormente. Muchos compuestos identificados correspondieron a compuestos previamente descritos en otras especies de *Solanaceae*, principalmente en tomate. Por tanto, en este trabajo, se identificó un total de 45 compuestos, 34 de los cuales son descritos en frutos de pimienta por primera vez.

La comparación cualitativa de los compuestos indicó algunas diferencias entre los extractos. De acuerdo con el número de compuestos identificados, el pimienta "dulce italiano" fue la variedad más rica, conteniendo 40 de todos los compuestos propuestos en este trabajo. En el caso de las otras dos variedades pimienta "Lamuyo" y pimienta "California wonder", se identificaron 33 y 35

¹⁹¹ Moiseev, D. V., Buzuk, G.N. & Shelyuto, V.L., 2011. Identification of flavonoids in plants by HPLC. *Pharmaceutical Chemistry Journal*, 45(1), pp.47–50.

¹⁹² Funari, C.S. *et al.*, 2011. Interconverting flavanone glucosides and other phenolic compounds in *Lippia salviaefolia* Cham. ethanol extracts. *Phytochemistry*, 72(16), pp.2052–2061.



compuestos, respectivamente. Sin embargo, seis compuestos no encontrados en la variedad verde se identificaron en las variedades roja y amarilla, siendo las dos últimas variedades más similares entre sí, con respecto a la variedad de pimiento verde.

1.3. Aplicación de la herramienta *clustering* de líneas y la herramienta de correlación a las muestras de pimiento para la caracterización y la identificación de compuestos bioactivos.

Se ha aplicado el método de *clustering* de líneas a los datos obtenidos de los análisis LC-MS de las tres variedades de pimiento para obtener los listados de posibles compuestos de las tres variedades de pimiento verde, amarillo y rojo. Una vez obtenidas las listas, se ha realizado la caracterización comparando con la caracterización convencional. Por último se ha realizado la correlación de los datos LC-MS con la capacidad antioxidante de los tres métodos utilizados, TEAC, FRAP y ORAC.

- **Caracterización: listado de *clusters*/compuestos.**

En la Figura 75 se muestra arriba los cromatogramas BPC y TIC y abajo, el espectrograma de baja resolución (1s, 1Da) de uno de los extractos analizados, el derivado de la variedad de pimiento verde. El espectrograma representa el logaritmo de la intensidad en función del tiempo (eje X) y la masa (eje Y) mediante una escala de color. Como se aprecia, es una muestra muy compleja con muchos compuestos, que presenta líneas en continuo y mucho ruido. En la Figura 76, se presentan dos detalles correspondientes a las zonas del espectrograma; a la izquierda la correspondiente al intervalo de tiempo [28.8min, 30min] y de masa m/z [429Da, 465Da] y a la derecha, un detalle donde aparece la masa $m/z=433$ Da. Claramente, se aprecian tres compuestos con sus variantes isotópicas.



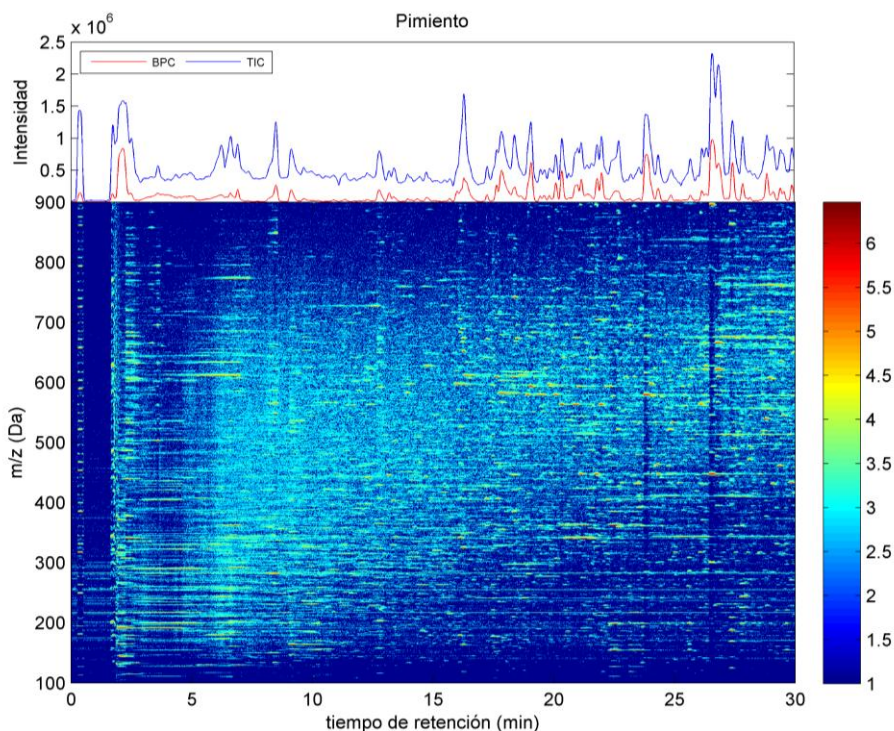


Figura 75. Arriba: cromatograms BPC y TIC de una muestra de pimienta verde. Abajo: Espectrograma de baja resolución (1 s,1 Da) de la misma muestra.

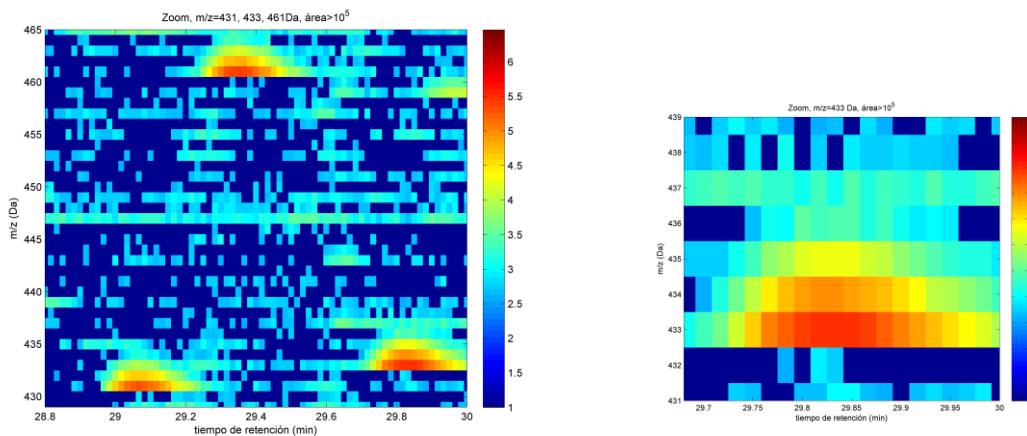


Figura 76. Izquierda: detalle de una zona del espectrograma de la Figura 75 para el intervalo de tiempo [28.8 min, 30 min] y de masa m/z [429 Da, 465 Da]. Derecha: detalle de una zona del espectrograma de la Figura 75 para el intervalo de tiempo [29.68 min, 30 min] y de masa m/z [431 Da, 439 Da].



Tabla 21. Datos de los ficheros de las muestras de pimienta.

Variedad	Ficheros .baf	Ficheros .bin	puntos iniciales($t_R, m/z, I$) del fichero .bin	puntos $I > 1000$ del fichero .bin
Verde	539.8 MB	25.6 MB	1.403.902	149.464
Amarillo	540.2 MB	26.0 MB	1.407.027	195.414
Rojo	529.6 MB	24.1 MB	1.367.886	175.062

En la Tabla 21, se presentan los datos referentes a los ficheros de las muestras de las tres variedades de pimienta. En primer lugar, es necesario aplicar un preprocesamiento a los datos LC-MS de todas las muestras, que consiste en la conversión de los ficheros a binario, la eliminación de ruido y la calibración en masa. Los parámetros utilizados en esta etapa previa se muestran en la Tabla 22, mientras que los parámetros utilizados para la realización del *clustering* de líneas se indican en la Tabla 23.

Tabla 22. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de pimienta.

Ficheros	pimiento: v1, a1, r1
Rango de tiempo	[0, 30 min] ó [0, 1800s]
Rango de masa	[100, 900 Da]
Eliminación ruido iónico y Umbral de intensidad	umbral_ruido=80 cuentas
Calibrante	formiato de Na

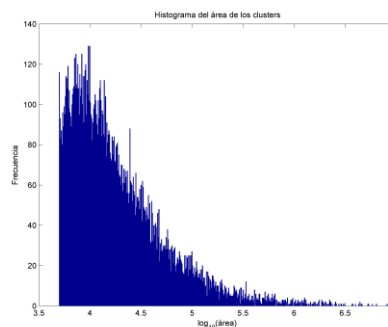


Tabla 23. Parámetros utilizados para el *clustering* de líneas de todas las muestras de pimiento analizadas.

Nombre del parámetro	Valor utilizado
Umbral intensidad	1000 cuentas
Umbral área	5000 cuentas
Elimina cluster, N_puntos	3
T_resol_fase1	300 s
T_resol_fase2	20 s
umbral_tr	20 s
umbral_m/z	0.010 Da



Clustering de líneas	Nº clusters
Pre-Clustering fase1:	7.531
Pre-Clustering fase2:	16.371
Elimina: puntos<3, área<5000	6.853
Clustering de puntos	14.984
Elimina: puntos<3, área<5000	7.573
Detección picos cromatogr.	13.045
Elimina: puntos<3, área<5000	11.231
Umbral de área > 10 ⁴	7.827
Umbral de área > 10 ⁵	948
Umbral de área > 10 ⁶	92



Tiempo computacional

Preclustering: fase 1 y 2	1.1 min	42%
Clustering de puntos	0.6 min	23.2%
Detección de picos cromatográficos	0.6 min	23.2%
Estadística	0.3 min	11.6%
TOTAL	2.6 min	100%

Figura 77. Izquierda: número de *clusters* generados en cada fase del *clustering* de líneas. Derecha: histograma del área de los *clusters* y tiempo computacional invertido para la muestra de pimienta verde.

En la Figura 77, se indica en la tabla de la izquierda el número de *clusters* generados en cada fase del *clustering*, y a la derecha el histograma del área y el tiempo computacional invertido durante el *clustering*. Como se aprecia en dicha figura, dependiendo del umbral de área que se establezca, se obtiene una lista con un número mayor o menor de *clusters*. Se ha generado una lista con 948 *clusters* (con área > 10⁵ cuentas), de los cuales 92 *clusters* tienen un área mayor de 10⁶ cuentas. Establecer un umbral de área adecuado es muy importante, ya que condiciona el número de *clusters* y por tanto, de compuestos que van a



poder ser identificados para la caracterización o que van a ser utilizados para posteriores estudios estadísticos.

En la Figura 78 se han representado los *clusters* que tiene un valor de área $>10^5$ cuentas. En la Figura 79 se muestran los *clusters* correspondientes a un detalle de las figuras anteriores. Se comprueba como el método de *clustering* ha detectado los tres compuestos.

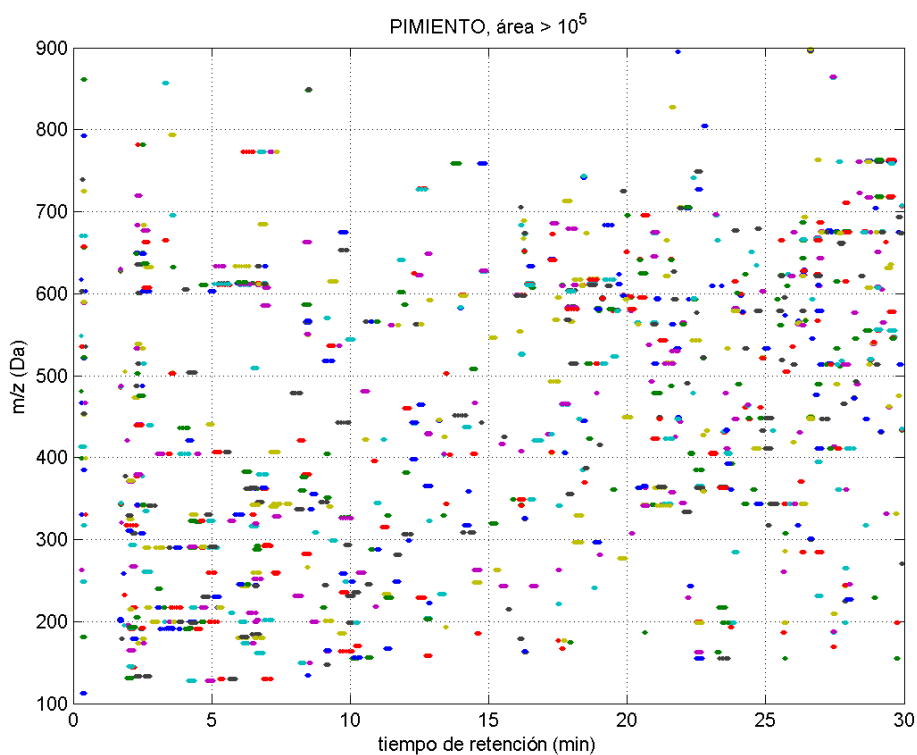


Figura 78. Representación de los *clusters* generados mediante la técnica de *clustering* lineal para la muestra de pimiento verde.



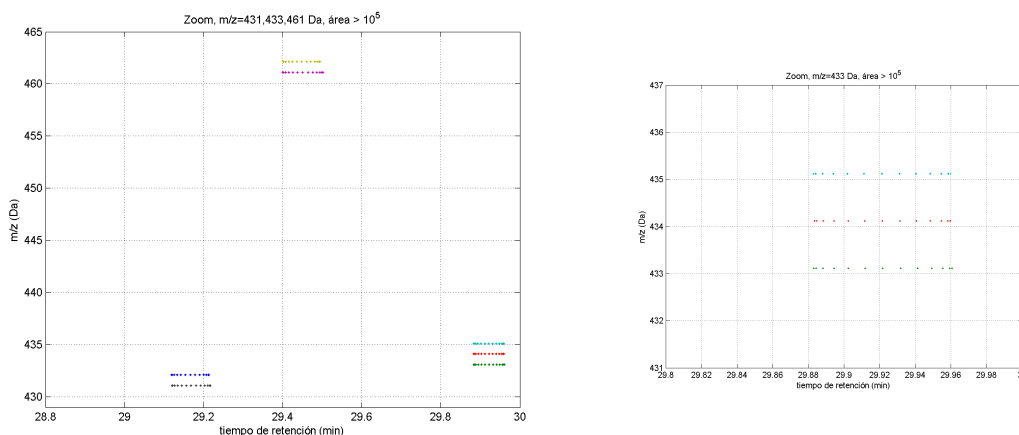


Figura 79. Representación de los *clusters* generados mediante la técnica de *clustering* lineal para el detalle de la zona de la Figura 78.

En la Tabla 24 se presenta los parámetros de los *clusters* del listado correspondientes a la zona estudiada: el parámetro Num., que indica la posición que ocupa en orden de área decreciente, el área en Kcuentas, la masa y su desviación estándar en Da y mDa, respectivamente, la máxima diferencia entre las masas de los puntos que componen los *clusters* (rangoM en mDa), el tiempo de retención y su desviación estándar en segundos y la diferencia máxima de los t_R de los puntos que componen el *cluster* (rangoT en s). En el apéndice 1 se presenta una tabla con los primeros 100 *clusters* de mayor área identificados mediante este procedimiento.

Tabla 24. Parte del listado de compuestos resultados del *clustering* de líneas donde aparecen los tres compuestos del ejemplo.

Num.	Área (Kcuentas)	Masa (Da)	SE_m (mDa)	rangoM (mDa)	tr (s)	SE_tr (s)	Rango_tr (s)
3	633.2720	431.0984	0.6734	8.5754	29.205	0.034	0.217
5	138.8080	432.1015	1.3130	9.2773	29.202	0.030	0.150
41	2039.9680	433.1149	0.5108	3.7842	29.922	0.039	0.201
42	561.6040	434.1174	0.6134	3.5095	29.922	0.038	0.167
13	103.0400	435.1198	0.9027	3.4180	29.923	0.036	0.167
38	1777.6440	461.1090	0.5065	13.4583	29.450	0.048	0.452
39	393.3960	462.1116	0.8246	4.5166	29.447	0.045	0.251



Se ha aplicado el algoritmo de detección de huellas espectrométricas para identificar los compuestos y se han identificado 164 compuestos con área mayor de 200 Kcuentas, de los cuales 45 coinciden con los que se habían identificado con el método convencional (que se presentan en la Tabla 19). En el apéndice 1, se muestra una tabla con los *clusters* eliminados porque salen en continuo, los 164 compuestos y sus familias completas (fragmentos, isótopos,...) o huellas espectrométricas de los compuestos identificados (que aparecen marcados en color gris).

En la Figura 80 se representa el cromatograma BPC (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área mayor de 2×10^5 cuentas (en rojo). En total, son 164 picos representantes (o familias de líneas). Mediante el *clustering*, se han identificado todos los mayoritarios, y además muchos minoritarios (de poca intensidad) que aparecen ocultos debajo de los mayoritarios. Estos picos ocultos sería imposible identificarlos mediante el método convencional. Los picos minoritarios que coeluyen con los mayoritarios y que permanecen ocultos se pueden apreciar mejor en la Figura 81, donde se muestran detalles de diferentes zonas del cromatograma.

En la Figura 82 se representa en azul el BPC y en rojo sólo los picos que cumplen tener una intensidad superior a 10^5 cuentas. Se observa que todos los picos que cumplen esta condición (59 picos) ha sido posible identificarlos mediante el *clustering* de líneas. Al ser picos mayoritarios en la mayoría de los casos es posible identificarlos mediante el método convencional consistente en la detección de los picos del cromatograma. Pero algunos picos es posible que no puedan detectarse (aun siendo mayoritarios) si quedan ocultos bajo otros picos de mayor intensidad, como se aprecia en la Figura 83. A la izquierda, se observa



un pico oculto de intensidad mayor de 10^5 y a la derecha todos los picos ocultos encontrados mediante el *clustering*.

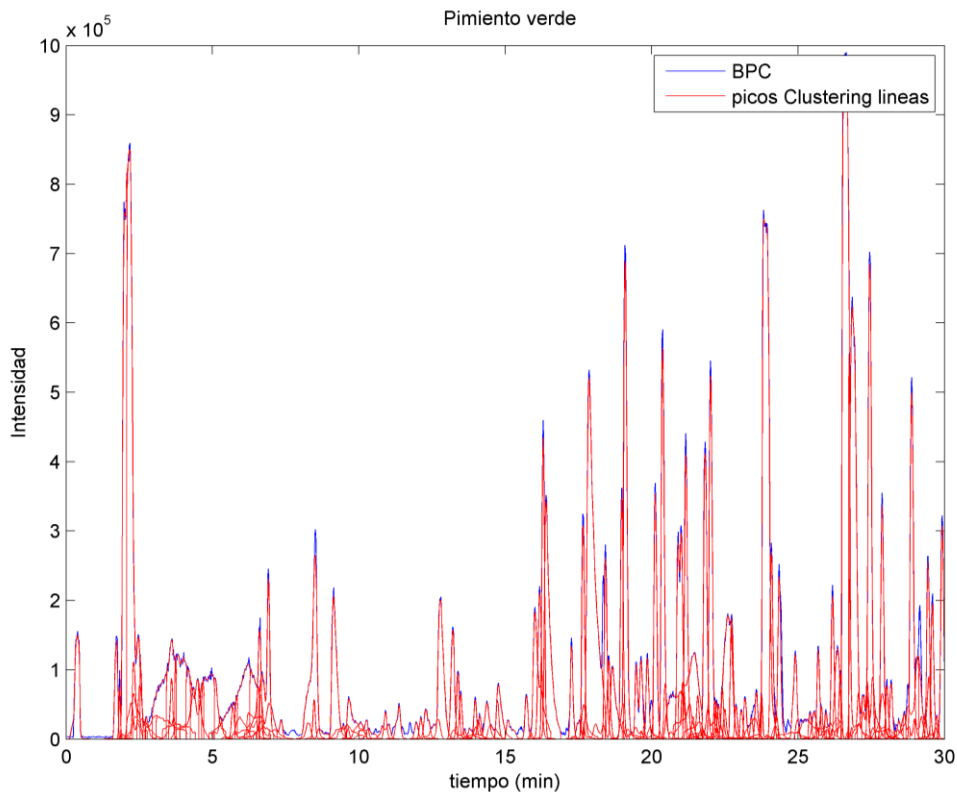
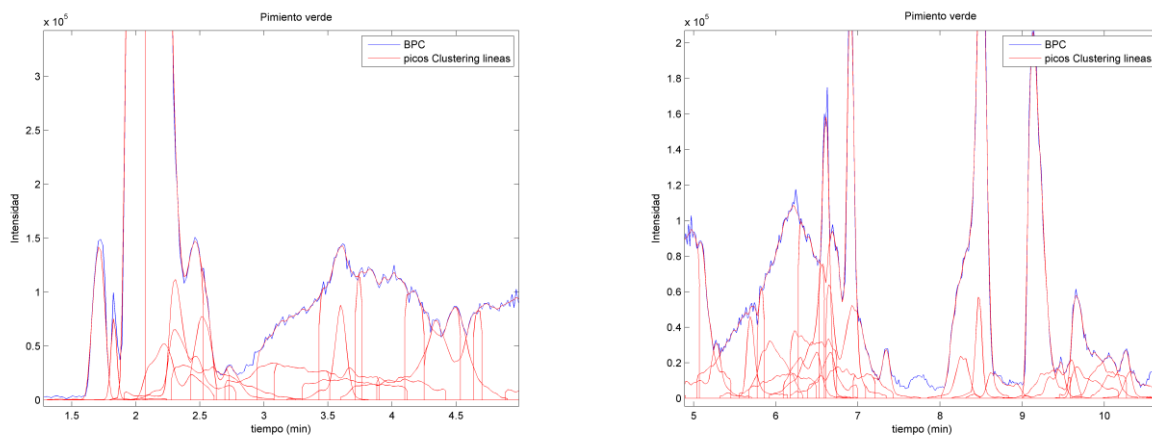


Figura 80. Cromatograma BPC del pimiento verde (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas (en rojo).



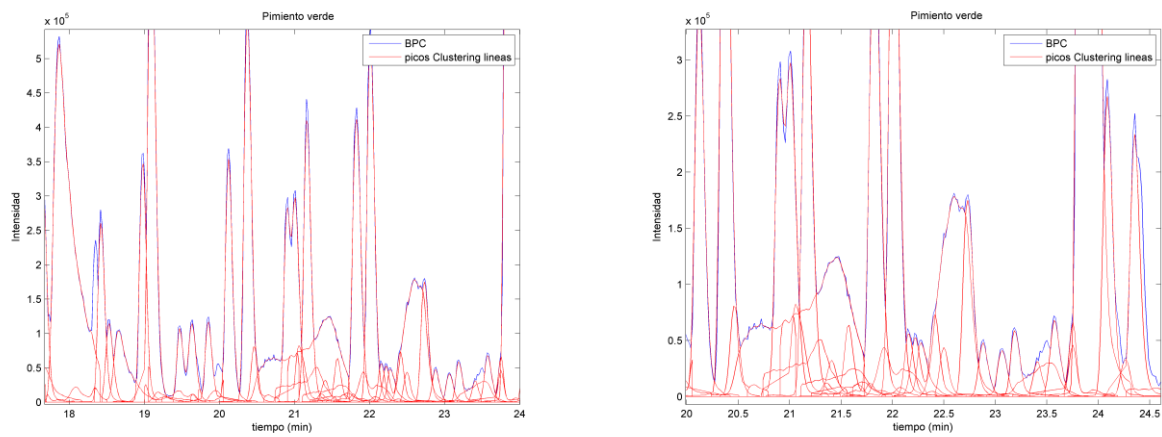


Figura 81. Diferentes detalles de la Figura 80. BPC (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas (en rojo).

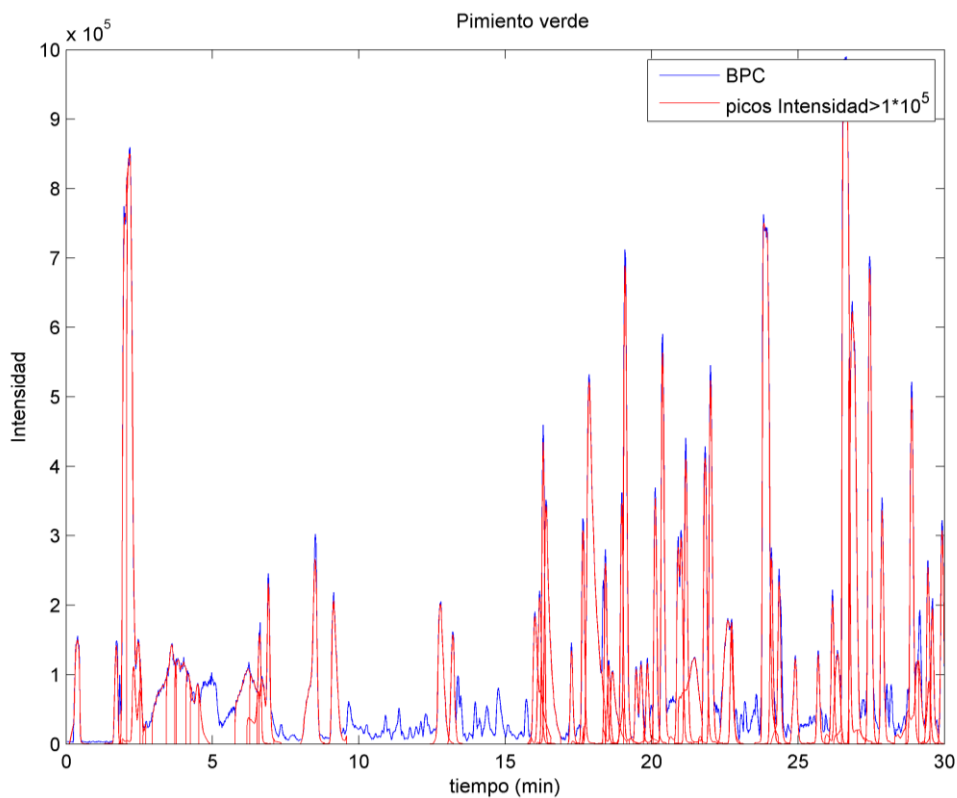


Figura 82. Cromatograma BPC del pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado sólo los picos con intensidad $> 10^5$ cuentas (59 picos en rojo).



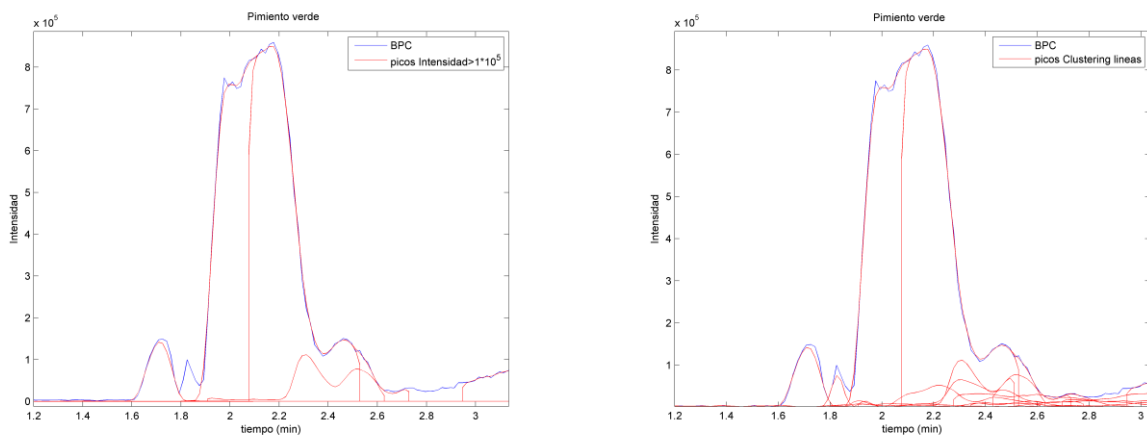


Figura 83. Izquierda: detalle de la Figura 82 correspondiente a un pico con intensidad superior a 10^5 cuentas oculto bajo otro pico de mayor intensidad. Derecha: mismo detalle pero representando todos los picos encontrados por el *clustering* (detalle de la Figura 80).

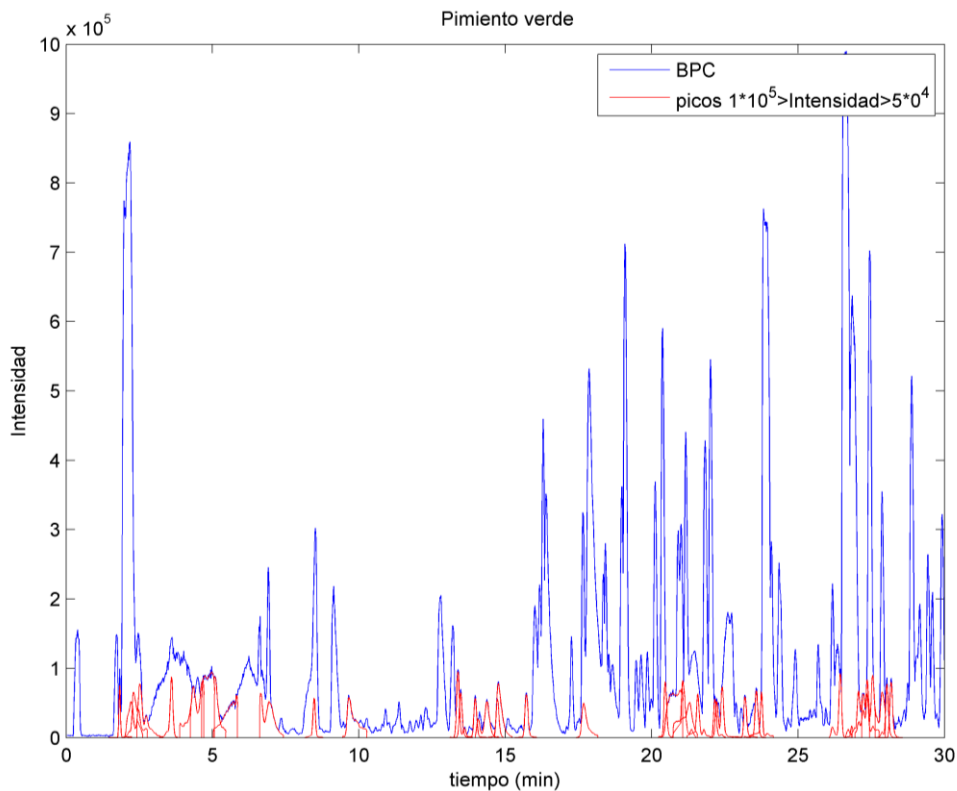


Figura 84. Cromatograma BPC del extracto de pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado sólo los picos con intensidad comprendida entre 10^5 y 5×10^4 cuentas (37 picos en rojo).



Mediante el método de *clustering* aplicado se han podido detectar muchos picos ocultos con intensidades menores como muestran la Figura 84 (37 picos con intensidad entre 10^5 y 5×10^4 cuentas) y la Figura 85 (68 picos con intensidad menor de 5×10^4 cuentas)

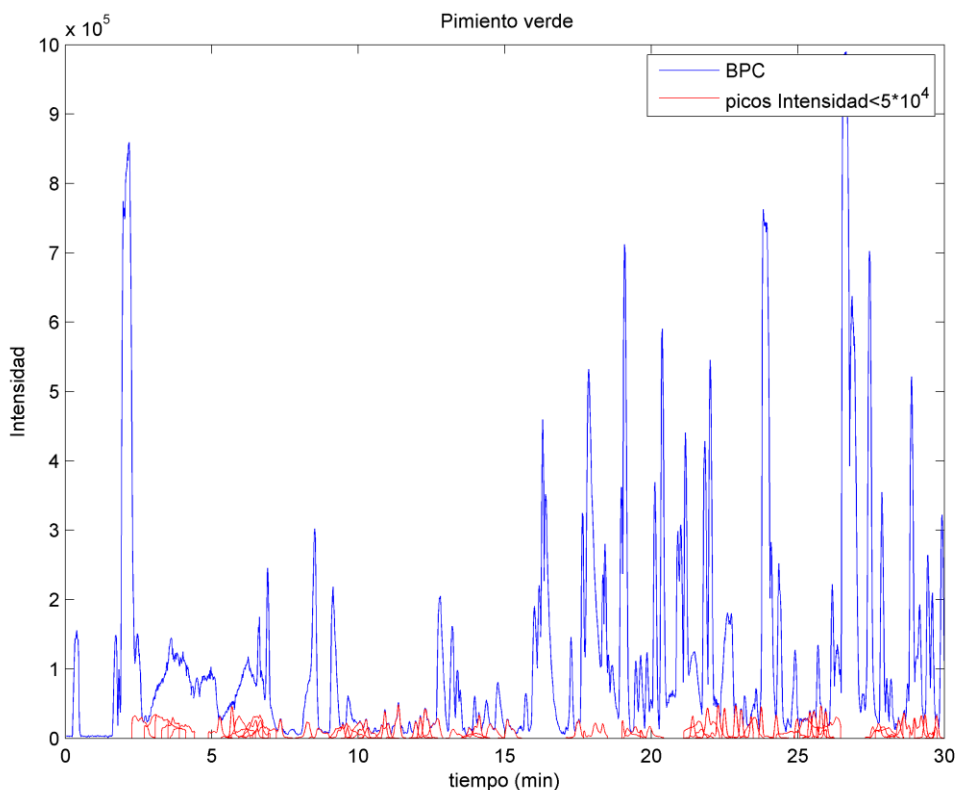


Figura 85. Cromatograma BPC del pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado los picos con intensidad $< 5 \times 10^4$ cuentas (68 picos en rojo).

Como se aprecia el número de picos encontrados mediante el método de *clustering* de líneas es elevado y depende de los umbrales de área establecidos. El número de picos detectados es muy superior al que es posible detectar mediante el método convencional, por lo que la herramienta de *clustering* se presenta como una herramienta con gran potencial para la caracterización de muestras complejas.



- **Correlación entre los datos LC-MS y capacidad antioxidante.**

Posteriormente, se realizó el estudio de correlación entre los datos LC-MS de los extractos de las tres variedades de pimiento y los datos de capacidad antioxidante. La correlación se estudió de manera separada según el método de capacidad antioxidante utilizado: TEAC, FRAP y ORAC.

El modelo de correlación consiste en un modelo de regresión lineal simple entre el área de cada *cluster* y la bioactividad de cada muestra. En este estudio, se dispone de pocas muestras de pimiento respecto al gran número de áreas de los compuestos encontrados, por lo que se aplicó una correlación mediante un función lineal, tal y como se especifica en la metodología (Ecuación 11).

Como resultado del análisis de correlación, se obtienen los coeficientes de correlación para cada *cluster*. El algoritmo proporciona de forma automática una lista de candidatos a compuestos bioactivos, ordenados según el valor de r . El programa proporciona los candidatos con la información necesaria para su identificación (t_R , m/z , desviaciones estándar), además de las áreas calculadas para cada candidato en cada muestra.



Tabla 25. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método TEAC de las muestras de pimienta.

#	r	m/z (Da)	t _R (min)	A (verde)	A (amarillo)	A (rojo)
1	0.9992	191.0205	4.10	5135.8	7770.1	2317.1
2	0.9991	355.1028	17.95	889.5	1296.1	363.9
3	0.9965	361.1497	19.54	773.9	1085.4	293.6
4	0.9866	175.0387	17.95	350.5	457.4	120.2
5	0.9852	243.1711	15.54	197.5	302.8	113.8
6	0.9840	186.1144	20.63	202.9	331.2	117.4
7	0.9801	385.1866	18.39	592.8	1028.6	324.0
8	0.9734	218.1037	10.91	509.6	648.5	257.3
9	0.9527	443.1924	14.74	1012.7	1144.1	370.9
10	0.9457	361.1507	24.55	179.3	350.5	111.8
11	0.9434	425.2026	21.73	220.3	260.4	115.1
12	0.9148	201.1250	9.11	319.2	456.3	234.8
13	0.9011	445.2065	18.45	868.8	1855.7	445.8
14	0.8822	243.1712	16.70	254.4	259.7	122.0
15	0.8768	573.1855	29.54	442.9	799.2	100.0
16	0.8734	536.1140	6.45	288.2	512.5	261.0
17	0.8711	259.1290	10.35	510.1	498.4	213.5
18	0.8710	131.0471	2.08	310.1	300.3	100.9
19	0.8562	435.9577	4.05	211.4	257.9	152.0
20	0.8468	493.2285	28.27	515.6	1089.4	173.4
21	0.8182	215.0328	2.11	171.0	285.3	166.7
22	0.8109	191.0207	4.62	2704.2	2505.9	1292.7
23	0.8048	411.1328	28.03	452.2	1169.9	366.5

r: coeficiente de correlación

t_R: tiempo de retención

A: área (en Kcuentas=1000cuentas)



Tabla 26. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método FRAP de las muestras de pimienta.

#	r	m/z (Da)	t _R (min)	A (verde)	A (amarillo)	A (rojo)
1	0.9932	218.1037	10.91	5135.8	7770.1	2317.1
2	0.9845	175.0387	17.95	889.5	1296.1	363.9
3	0.9840	361.1497	19.54	350.5	457.4	120.2
4	0.9817	355.1028	17.95	202.9	331.2	117.4
5	0.9808	443.1924	14.74	509.6	648.5	257.3
6	0.9777	191.0205	4.10	1012.7	1144.1	370.9
7	0.9762	425.2026	21.73	220.3	260.4	115.1
8	0.9712	243.1711	15.54	197.5	302.8	113.8
9	0.9547	186.1144	20.63	319.2	456.3	234.8
10	0.9342	243.1712	16.70	868.8	1855.7	445.8
11	0.9307	385.1866	18.39	254.4	259.7	122.0
12	0.9253	259.1290	10.35	442.9	799.2	100.0
13	0.9175	201.1250	9.11	288.2	512.5	261.0
14	0.9166	131.0471	2.08	510.1	498.4	213.5
15	0.8841	435.9577	4.05	310.1	300.3	100.9
16	0.8753	191.0207	4.62	211.4	257.9	152.0
17	0.8633	361.1507	24.50	515.6	1089.4	173.4
18	0.8256	536.1140	9.45	171.0	285.3	166.7
19	0.8146	325.1249	2.10	2704.2	2505.9	1292.7
20	0.7791	215.0328	2.11	452.2	1169.9	366.5
21	0.7775	259.1288	7.33	147.9	197.9	132.1

r: coeficiente de correlación

t_R: tiempo de retención

A: área (en Kcuentas=1000cuentas)



Tabla 27. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método ORAC de las muestras de pimiento.

#	r	m/z (Da)	t _R (min)	A (verde)	A (amarillo)	A (rojo)
1	0.9488	145.0631	2.14	1718.7	1255.0	1696.9
2	0.9050	342.1173	6.58	581.0	281.6	525.5
3	0.8886	232.8982	1.86	217.4	142.3	173.8
4	0.8647	202.1079	6.66	255.8	125.9	203.2
5	0.8135	292.1389	6.66	783.3	380.0	597.8
6	0.8130	130.0884	5.29	159.5	105.6	121.5
7	0.7947	521.2024	24.90	1166.2	510.7	1087.3

r: coeficiente de correlación

t_R: tiempo de retención

A: área (en Kcuentas=1000cuentas)

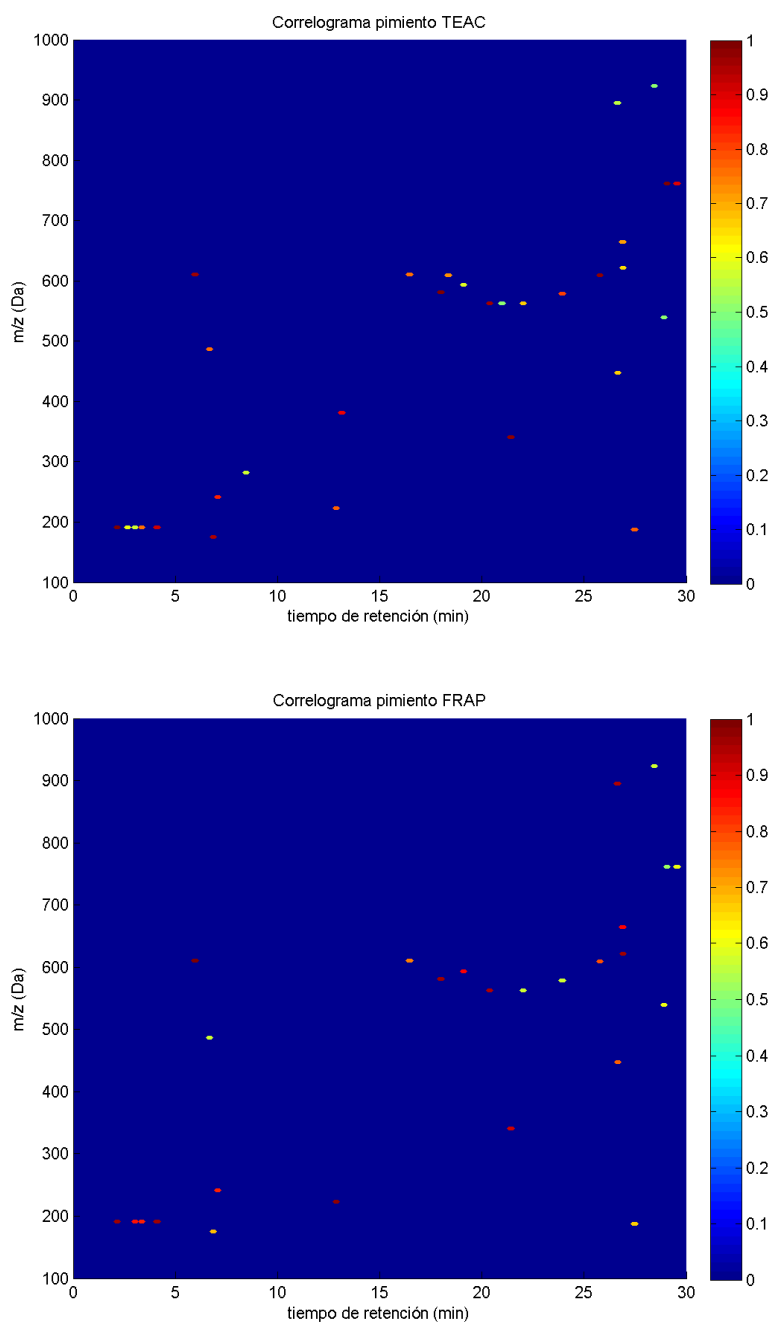
En la Tabla 25, la Tabla 26 y la Tabla 27 se presentan los compuestos candidatos ordenados por coeficiente de correlación (*r*). Se indican los valores de masa y t_R de los *clusters* con los coeficientes de correlación obtenidos para TEAC, FRAP y ORAC, respectivamente. Dichos coeficientes indican que el compuesto correspondiente a dicho *cluster* está estadísticamente asociado con la bioactividad. Por lo tanto, los compuestos presentados en estas tablas son los candidatos a compuestos bioactivos. Mediante los parámetros establecidos al aplicar la herramienta de correlación ($r > 0.7$), se obtuvieron 23 candidatos para el TEAC y 21 candidatos para el FRAP, mientras que la lista de candidatos de ORAC fue de 7 compuestos candidatos.

Los compuestos potencialmente bioactivos obtenidos con los métodos TEAC y FRAP son muy similares y totalmente distintos de los obtenidos con el método ORAC. Los coeficientes de correlación son además más elevados al correlacionar con los datos TEAC y FRAP, con respecto a la utilización de los datos ORAC.

Una forma de representar los coeficientes de correlación gráficamente es mediante lo que se denomina correlograma o representación del coeficiente de



correlación (r) mediante un mapa de color en función del tiempo y la masa de cada *cluster*. En la Figura 86, se muestran los correlogramas correspondientes a los estudios de correlación con los tres métodos de capacidad antioxidante TEAC, FRAP, y ORAC. En los correlogramas, los compuestos candidatos a ser los responsables de la capacidad antioxidante se observan en color rojo (corresponden con coeficientes de correlación próximos a 1).



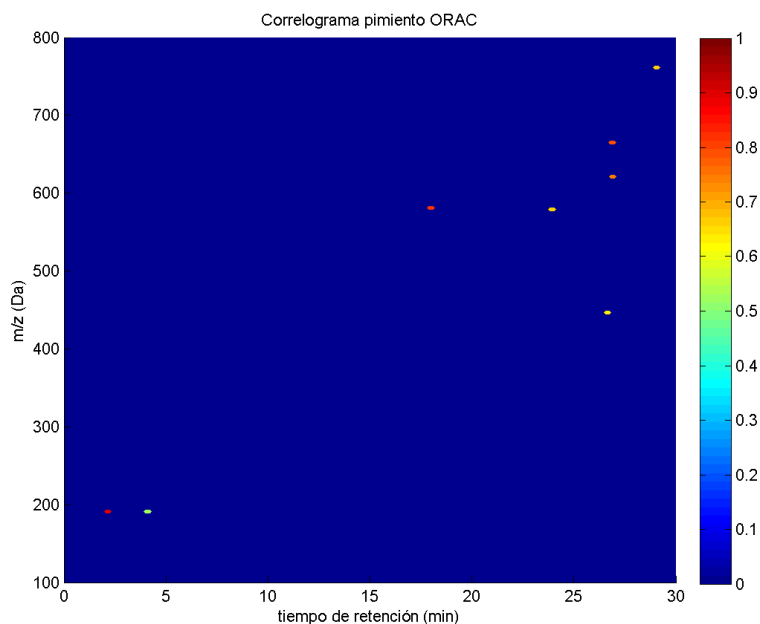


Figura 86. Correlogramas correspondientes a los estudios de correlación entre los datos LC-MS y la capacidad antioxidante de los tres métodos TEAC, FRAP, y ORAC para las muestras de pimiento.

La similitud en cuanto a los resultados de correlación entre los métodos TEAC y FRAP, así como las diferencias encontradas entre el método ORAC puede deberse al mecanismo de acción en el que se basan dichos métodos. Los métodos de capacidad antioxidante TEAC y FRAP se basan en la transferencia de electrones, mientras que el mecanismo del método ORAC se basa en la transferencia de átomos de hidrógeno.

Convencionalmente, para estudiar la relación entre la composición de extractos vegetales y su capacidad antioxidante de una manera práctica, se suele correlacionar el contenido fenólico total con la capacidad antioxidante. Sin embargo, esta forma no permite identificar el/los compuesto/s potencialmente responsables de la capacidad antioxidante observada. Para atribuir la capacidad antioxidante a ciertos tipos de compuestos se realiza el estudio de los componentes mayoritarios de los perfiles cromatográficos y se estudia su

estructura mediante lo que se denomina estudio de la relación estructura-actividad (SAR). Según los estudios SAR referentes a los compuestos fenólicos, los compuestos responsables de la capacidad antioxidante son los pertenecientes a las familias de los ácidos fenólicos y los flavonoides. En ambos casos, está muy relacionada con el número y la posición de los grupos hidroxilo (-OH)^{193,194}.

A continuación, se discuten los resultados obtenidos, comenzando por una comparación entre los candidatos de TEAC y FRAP y terminando por la discusión de los candidatos de ORAC.

Dentro de los ácidos fenólicos, se sabe que los ácidos hidroxicinámicos están relacionados con una mayor capacidad antioxidante que los ácidos hidroxibenzóicos, debido al doble enlace que poseen en su estructura. Esto concuerda con los resultados de nuestro estudio, puesto que el candidato de m/z 355.1028 Da a 17.95min, que aparece tanto en la lista de TEAC como en la lista de FRAP con coeficientes de correlación elevados ($r=0.9991$ y $r=0.9817$, respectivamente) es el feruloil glucósido, un ácido hidroxicinámico identificado en pimienta. Se observa que el candidato de m/z 175.0387 Da eluye al mismo tiempo de retención y se ha identificado como fragmento del feruloil glucósido anteriormente citado. Para dicho candidato, se obtuvo un coeficiente de correlación de $r=0.9866$ para TEAC y $r=0.9845$ para FRAP. El hecho de obtener alta correlación para el compuesto y su fragmento corrobora que este ácido hidroxicinámico es un buen candidato responsable de la capacidad antioxidante.

También aparece como compuesto candidato para ambos métodos el compuesto detectado a 218.1037 Da y 10.91 min, que corresponde con el ácido

¹⁹³ Wojdylo, a, Oszmianski, J. & Czemerys, R., 2007. Antioxidant activity and phenolic compounds in 32 selected herbs. *Food Chemistry*, 105(3), pp.940–949.

¹⁹⁴ Sekher Pannala, A. *et al.*, 2001. Flavonoid B-ring chemistry and antioxidant activity: fast reaction kinetics. *Biochemical and Biophysical Research Communications*, 282(5), pp.1161–1168.



pantoténico, también conocido como vitamina B5. Se trata del primer candidato de la lista de FRAP (con $r=0.9932$) y el octavo candidato de la lista de TEAC (con $r=0.9734$).

Encabezando la lista de candidatos de TEAC ($r= 0.9992$) aparece un compuesto de m/z 191.0205 a 4.10 min, el cual se ha identificado como ácido cítrico. Este mismo compuesto también aparece como candidato en la lista de FRAP, aunque con una correlación ligeramente menor ($r= 0.9777$). Además, se ha obtenido un isómero del ácido cítrico que eluye a 4.62 que también aparece como candidato en TEAC y FRAP. Otro compuesto con alta correlación es el detectado a 131.0471 Da y 2.10 min, que probablemente corresponda con un fragmento del ácido quínico. La correlación de este candidato fue mayor con los datos FRAP ($r= 0.9166$) que con los datos TEAC ($r= 0.8710$). Tanto el ácido cítrico como el quínico son un ácidos orgánicos muy polares y la alta correlación obtenida para puede deberse a su alta intensidad en todas las muestras analizadas, sobre todo en el caso del ácido quínico.

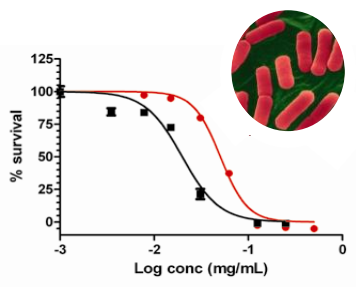
Entre los candidatos obtenidos al correlacionar los datos TEAC, el compuesto de m/z 445.2065 Da que eluye a 18.45 min se ha identificado como fragmento de la quercetina dihexósido. Se ha de destacar este candidato puesto su alta correlación ($r=0.9011$) indica que la relación estadística con la bioactividad probablemente dependa de la concentración del ión precursor, un flavonol de m/z 609,1504 Da ($C_{27}H_{30}O_{16}$). Además, este compuesto se propone como candidato exclusivamente en el método de capacidad antioxidante TEAC.

Los candidatos proporcionados al correlacionar con los valores ORAC, poseen un coeficiente de correlación menor y la gran mayoría poseen masa par. Entre ellos, destaca el compuesto detectado a m/z 145.0631 Da y 2.14 min, que posiblemente sea un fragmento del ácido quínico ($r=0.9488$). Otro de los compuestos que ha proporcionado una correlación relativamente alta



($r=0.8647$) es el detectado a m/z 202.1079 Da y 6.66 min. Este compuesto de masa par y fórmula molecular $C_9H_{16}NO_4$ se ha identificado tentativamente como un fragmento del nucleósido uridina (243,0634 Da, $C_9H_{11}N_2O_6$). Por último, se ha de destacar entre los candidatos del método ORAC, el compuesto de m/z 521.2024 Da a 24.90 min. Dicho candidato corresponde con un lignano glucosilado, el lariciresinol glucopiranosido.





BLOQUE II

Caracterización e identificación de compuestos antimicrobianos en muestras de *Cistus*

BLOQUE 2. CARACTERIZACIÓN E IDENTIFICACIÓN DE COMPUESTOS ANTIMICROBIANOS EN MUESTRAS DE *CISTUS*.

En el presente bloque se presentan los resultados del estudio de diferentes especies de la familia de las *Cistáceas* o familia *Cistaceae*. Esta familia de plantas, también conocida como jara, crece mayoritariamente en ecosistemas Mediterráneos semiáridos. Incluye el género *Cistus*, entre cuyas especies destacan *C. ladanifer*, *C. salviifolius*, *C. clusii* y *C. albidus*, por su abundancia en la Península Ibérica¹⁹⁵. Algunas de las especies de la familia *Cistaceae* son ricas en resinas inflamables, de manera que su crecimiento se asocia con problemas medioambientales e incendios forestales¹⁹⁶. En este sentido, los desechos de esta planta pueden considerarse una fuente rentable de ingredientes potencialmente bioactivos con aplicaciones industriales importantes. Ciertas especies son más apreciadas en la industria de la perfumería, mientras que otras se han utilizado desde la antigüedad en medicina tradicional. El antiguo uso etnobotánico de las especies de *Cistus* revela que estas plantas constituyen un buen remedio para algunas enfermedades e infecciones microbianas^{197,198}. Es bien conocida su actividad antimicrobiana, principalmente atribuida a la parte apolar, como por ejemplo, el aceite esencial de *C. salviifolius*¹⁹⁹ o extractos obtenidos mediante disolventes orgánicos de *C. creticus*²⁰⁰ y *C.*

¹⁹⁵ Stübing, G. & Peris, J.B., 1998. Plantas medicinales de la Comunidad Valenciana. Generalitat Valenciana, Conselleria de Medio Ambiente.

¹⁹⁶ Ferrandis, P., Herranz, M. & Mart, J.J., 1999. Effect of fire on hard-coated Cistaceae seed banks and its influence on techniques for quantifying seed banks. *Plant Ecology*, 144, pp.103–114.

¹⁹⁷ Bassolé, I.H.N. & Juliani, H.R., 2012. Essential oils in combination and their antimicrobial properties. *Molecules*, 17(4), pp.3989–4006.

¹⁹⁸ Salin, O.P. *et al.*, 2011. Effects of coadministration of natural polyphenols with doxycycline or calcium modulators on acute *Chlamydia pneumoniae* infection in vitro. *The Journal of Antibiotics*, 64(11), pp.747–752.

¹⁹⁹ Gertsch, J., 2011. Botanical Drugs, Synergy, and Network Pharmacology: Forth and Back to Intelligent Mixtures. *Planta Medica*, 77, pp.1086–1098.

²⁰⁰ Güvenç, A. *et al.*, 2005. Antimicrobiological studies on Turkish *Cistus* species. *Pharmaceutical Biology*, 43, pp.178–183.



*ladanifer*²⁰¹. También se ha atribuido una importante actividad antimicrobiana a la parte polar, como por ejemplo, los extractos acuosos derivados de *C. ladanifer* y *C. populifolius*⁵¹.

En cuanto a la capacidad antimicrobiana de las especies de *Cistus*, la parte más ampliamente estudiada es la fracción apolar, esencialmente constituida por compuestos de naturaleza terpénica, mientras que la parte polar se encuentra relativamente poco documentada en comparación con la fracción apolar. Por todo ello, en este bloque se han analizado ambas fracciones tanto la fracción apolar como la fracción polar de diferentes especies de *Cistus*, seleccionando la fracción polar para el estudio de la bioactividad. En la Figura 87, se muestran de manera esquemática las diferentes partes en las que consiste este bloque, siguiendo el esquema de la metodología general de la tesis. En primer lugar, se estudió la fracción apolar mediante el análisis de los compuestos volátiles extraídos mediante microextracción en fase sólida (SPME), utilizando cromatografía de gases acoplada a espectrometría de masas. En segundo lugar, se estudió la fracción polar de extractos de *Cistus*, incluyendo el estudio de compuestos fenólicos mediante cromatografía líquida acoplada a espectrometría de masas, el estudio de la capacidad antimicrobiana y el estudio de correlación entre los datos de espectrometría de masas y los datos bioactividad obtenidos.

²⁰¹ Ferreira, S. *et al.*, 2012. Screening of antimicrobial activity of *Cistus ladanifer* and *Arbutus unedo* extracts. *Natural Product Research*, 26(16), pp.1558–1560.



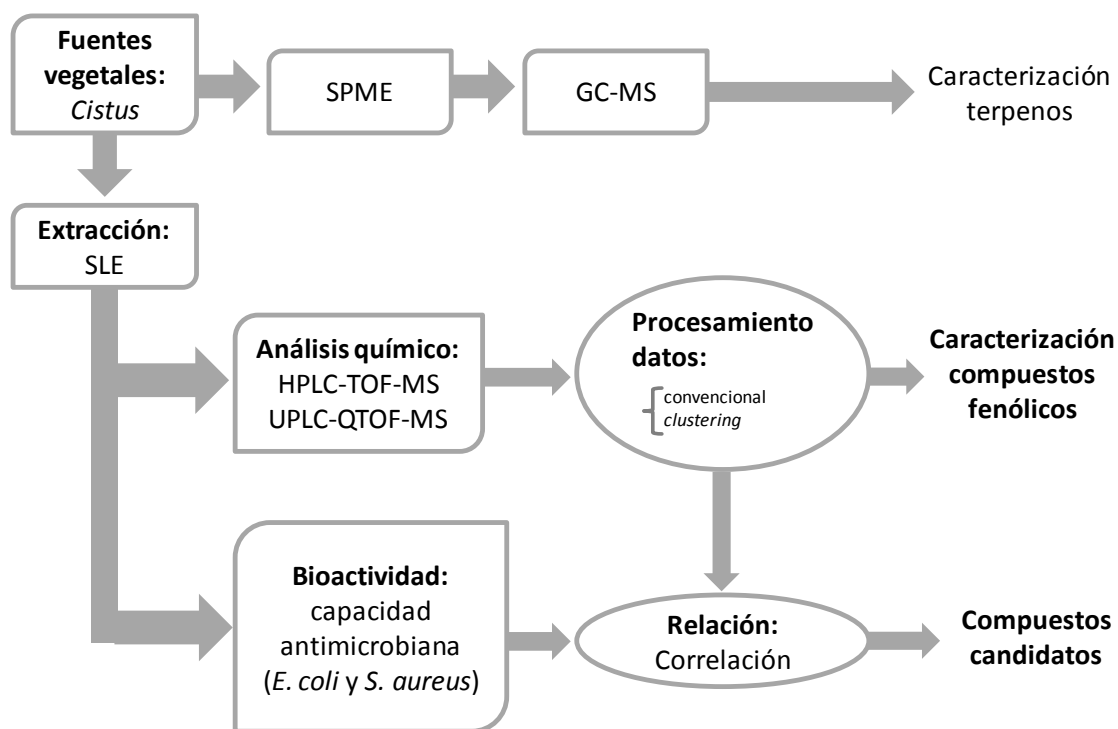


Figura 87. Esquema de los estudios realizados en el bloque 2.

A continuación, se presentan los resultados de los estudios de *Cistus* realizados, distinguiendo dos apartados referentes a la fracción apolar y la fracción polar, respectivamente.

2.1. Estudio de la fracción apolar de especies del género *Cistus*.

Las especies de *Cistus* incluidas en este estudio son las más representativas de éste género, con la peculiaridad de que pertenecen a distintos subgéneros: *Cistus ladanifer* y *Cistus salviifolius* corresponden al subgénero *Leucocistus*, *Cistus albidus* pertenece al subgénero *Cistus*, y *Cistus clusii* pertenece al subgénero *Halimioides*.

En la búsqueda de compuestos potencialmente bioactivos a partir de la fracción apolar de plantas del género *Cistus*, se decidió estudiar la fracción volátil. Para ello, se utilizó un equipo automatizado que permitió realizar en continuo la microextracción en fase sólida del espacio de cabeza y el análisis mediante



cromatografía de gases acoplada a espectrometría de masas de los compuestos previamente extraídos. De este modo, se obtuvieron los perfiles de cada una de las especies objeto de estudio. En los cromatogramas de iones totales (TIC, total ion chromatogram) obtenidos para *C. ladanifer*, *C. albidus*, *C. salviifolius* y *C. clusii* (Figura 88, Figura 89, Figura 90 y Figura 91, respectivamente), se incluyen las estructuras de algunos compuestos mayoritarios con la correspondiente intensidad a la que se han detectado. En la Tabla 28 se presentan los compuestos identificados, indicando el nombre común y el nombre según la nomenclatura IUPAC. Para cada compuesto identificado, se presenta el índice de retención de Kovats obtenido experimentalmente (Kovats index, KI), así como el porcentaje de área relativo a la suma de las áreas de la totalidad de los compuestos identificados en cada uno de los cuatro extractos analizados. Para algunos de los compuestos no se pudo comprobar el KI obtenido con los índices compilados en las bases de datos, con lo cual la identificación de éstos es tentativa, basándose únicamente en la información del espectro de masas.



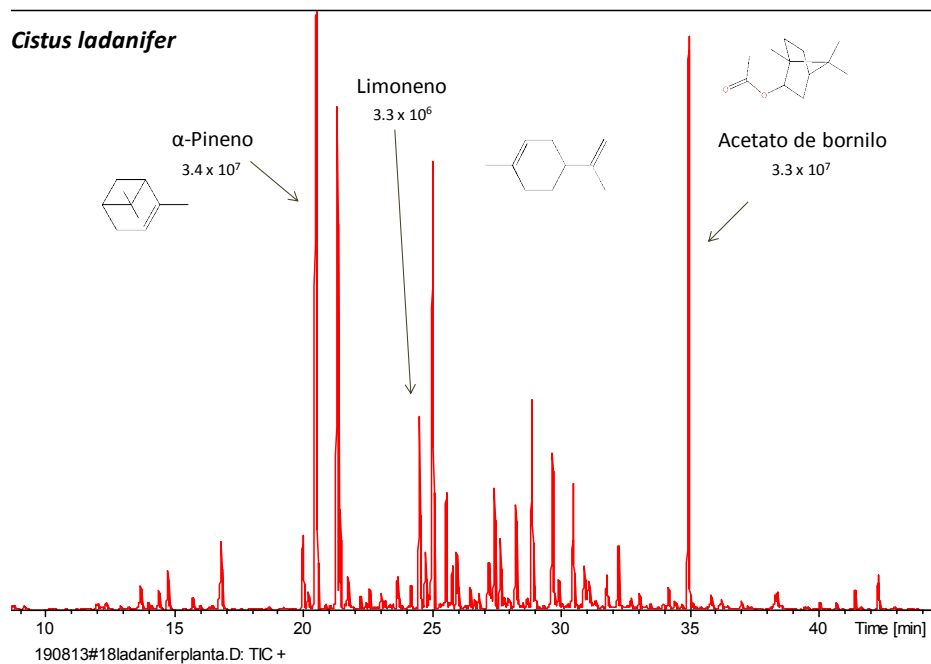


Figura 88. Cromatograma de iones totales de la muestra de *Cistus ladanifer* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.

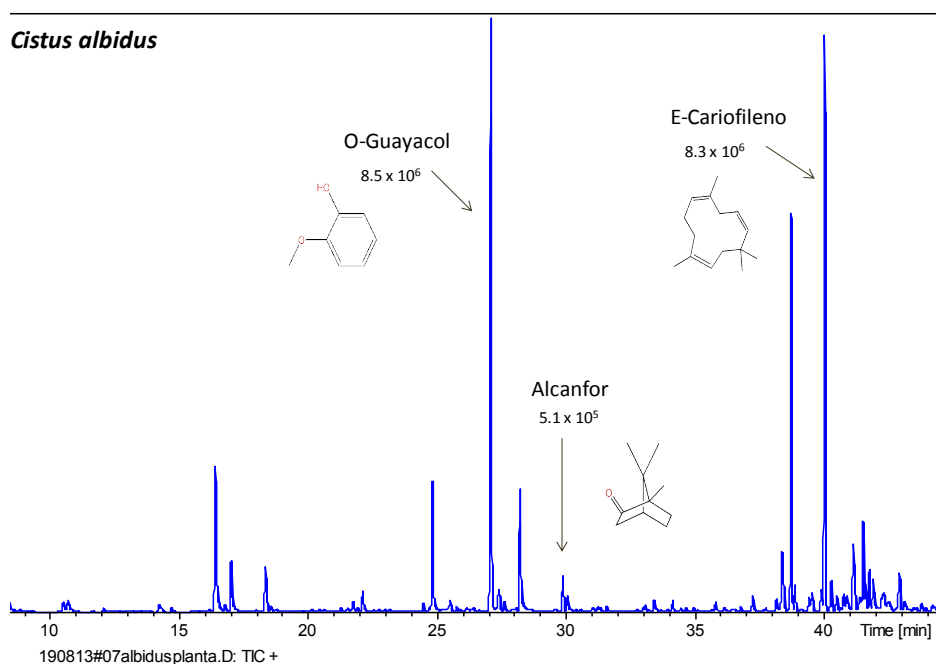


Figura 89. Cromatograma de iones totales de la muestra de *Cistus albidus* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.



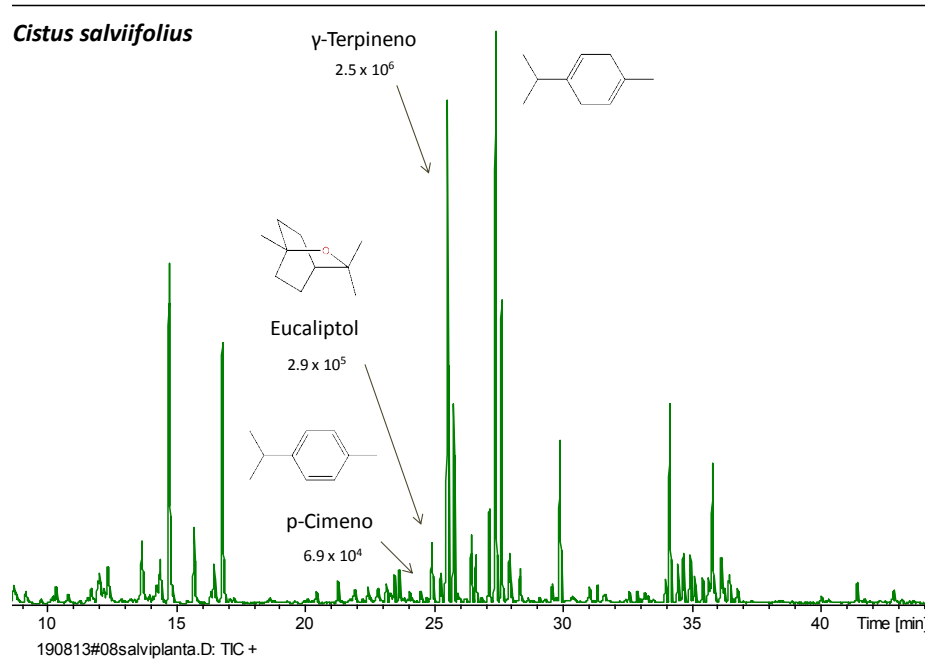


Figura 90. Cromatograma de iones totales de la muestra de *Cistus salviifolius* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.

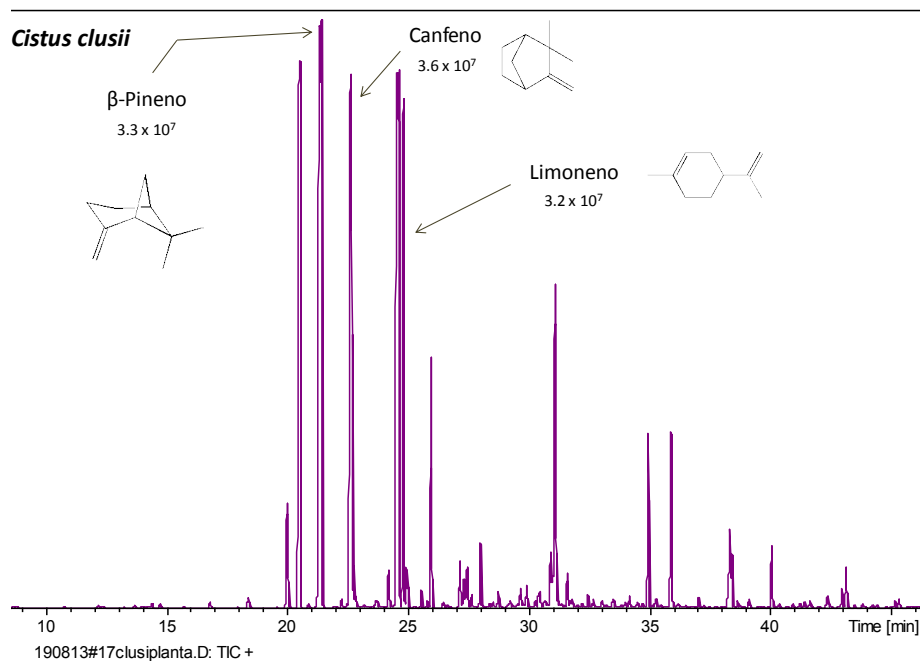


Figura 91. Cromatograma de iones totales de la muestra de *Cistus clusii* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.



Tabla 28. Compuestos identificados en las cuatro especies de *Cistus* analizadas mediante SPME-GC-MS.

#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
1	3-Pentanona ^①	700				0,0212
2	Pentanal	702	0,0407		0,0076	
3	2-Butanona, 3-hidroxi-	711	0,0534			
4	1-Butanol, 3-metil- ^①	737		0,0320		
5	1-Butanol, 2-metil- ^①	741		0,0456		
6	1-Pentanol	768		0,0069		
7	Benceno, metil- (Tolueno)	770	0,0449			
8	2-Butenal, 3-metil - ^①	784	0,0562			
9	1-Octeno ^①	791	0,0274			
10	2Z-Octeno ^①	806	0,0723		0,0045	0,0349
11	2E-Hexenal	853	0,0284		0,0113	
12	3-Hexen-1-ol, (3Z) -	855		0,4393		
13	n-Hexanol ^①	862 *		0,0190		
14	Benceno, etenil- (Estireno)	867		0,1544		0,1535
15	2-Heptanol ^①	894		0,1484		
16	Biciclo [3.1.0] hex-2-eno, 2-metil-5 -(1-metiletil) - (α -Tujeno)	929	0,8642			
17	Biciclo [3.1.1] hept-2-eno, 2, 6, 6-trimetil - (α -Pineno)	940	7,0422		0,0097	7,2097



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus Salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
18	2-Heptanona, 6-metil -	952	0,0420			
19	Biciclo [2.2.1] heptano, 2, 2-dimetil-3-metileno - (Canfeno)	957	5,3667		0,0189	11,7441
20	Biciclo [3.1.0] hex-2-eno, 4-metilen-1-(1- metiletil) - ^① (Tuja-2,4(10)-dieno)	960	0,9420			
21	4-hepten 1-ol, (4Z) - ^①	962		0,0114		
22	Benzaldehído	966	0,3654			
23	Biciclo [3.1.0] hexano, 4-metilen-1- (1-metiletil) - (Sabineno)	977	0,1345			0,0968
24	5-Hepten-2-ona, 6-metil-	981			0,0140	
25	Biciclo [3.1.1] heptano, 6, 6-dimetil-2-metileno- (β-Pineno)	984	0,2091			10,4895
26	1 ,6-Octadieno,7-metil-3-metileno - (Mirceno)	988				0,3970
27	Ciclopentanol, 3 - metil - ^①	1000 *			0,0122	
28	1,3-Ciclohexadieno, 2-metil-5- (1-metiletil) - (α-Felandreno)	1010				0,0701
29	1,3-Ciclohexadieno, 1-metil-4-(1-metiletil) - (α-Terpineno)	1020	0,2459			0,4104



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
30	Benceno, 1-metil-4-(1-metiletil) - (p-Cimeno)	1028	1,8345	0,0203	0,0115	
31	Ciclohexeno, 1-metil-4-(1-metiletenil) - (Limoneno)	1034	0,6162	0,0091		5,5023
32	Bencenometanol (Alcohol bencílico)	1035		0,3397		
33	Ciclohexeno, 3-metileno-6-(1-metiletil) - (β-Felandreno)	1037	0,1462			
34	2-Oxabiciclo [2.2.2]octano, 1,3,3 -trimetil - (Eucaliptol)	1037			0,0603	0,2027
35	Ciclohexanona, 2,2,6-trimetil-	1040	5,0811			
36	Fenol, 2-metil - (o-Cresol)	1050		0,0074		
37	1,4-Ciclohexadieno, 1-metil-4-(1-metiletil) - (γ-Terpineno)	1062	0,5385		0,0048	2,3094
38	2-Ciclohexen-1-ona, 3-metil-	1063 *	0,1924			
39	Etanona, 1-fenil - (Acetofenona)	1070	0,0815			
40	Fenol , 2-metoxi - (o-Guayacol)	1088		1,3704		
42	Benceno, 1-metil-3-(1-metiletenil) - ^① (p-Cimeneno)	1093	0,1320	0,0076		0,3256



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus Salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
43	1,6-Octadien-3-ol,3,7- dimetil- (Linalol)	1097				0,6681
44	2H-Piran , tetrahidro-4-metil-2-(2-metil-1-propen-1-il)-, (2S, 4R) - (Óxido de rosa (cis-))	1110				0,6494
45	Bencenoetanol (Fenil etil alcohol)	1116	1,1700	0,2971		
46	2H-Piran, tetrahidro-4-metil-2-(2-metil-1-propen-1-il)-, (2R, 4R) - (Óxido de rosa (trans-))	1128				0,1649
47	3-Ciclopenteno-1-acetaldehído, 2,2,3-trimetil-, (1R) - (α -Camfolenal)	1132	1,9287			
48	3-Nonen-2-ona (trans-)	1135	0,2404			
49	Benceno, 1,2-dimetoxi- ^① (Veratrol)	1141				0,1084
50	Biciclo [3.1.1] heptan-2-uno, 6,6-dimetil - ^① (Nopinona)	1149				0,0372
51	Biciclo [3.1.0] hexano-3-ol, 4-metilen-1-(1-metiletil)-, (1S, 3R, 5S) - ^① (Sabinol (trans-))	1152				0,2692
52	Biciclo [3.1.1] heptan-3-ol, 6, 6-dimetil-2-metileno-, (1R , 3S, 5R)-rel - (Pinocarveol (trans-))	1152	1,8883			
53	Biciclo [2.2.1] heptan-2-ona, 1,7,7-trimetil - (Alcanfor)	1158	0,3692	0,0853	0,1382	0,2612
54	Fenol, 4-etil-	1162		0,0452		



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
55	Biciclo [2.2.1] heptan-2-ol, 2,3,3-trimetil - (Canfeno hidrato)	1167	0,0681			0,0805
56	Biciclo [3.1.1] heptan-3-ona, 2,6,6-trimetil-, (1R, 2S, 5S)-rel - (Pinocanfona (trans-))	1171				0,1531
57	Biciclo [3.1.1] heptan-3-ona, 6,6-dimetil-2-metileno - (Pinocarvona)	1172	1,2631			0,1657
58	Biciclo [2.2.1] heptan-2-ol, 1,7,7-trimetil-, (1R, 2S, 4R)-rel - (Borneol)	1183	0,5824			
59	3-Ciclohexen 1-ol, 4-metil-1-(1-metiletenil)- ^① (1,8-Mentadien-4-ol)	1185				0,2492
60	3-Ciclohexen 1-ol, 4-metil-1-(1-metiletil)- (Terpinen-4-ol)	1187	0,5025			
61	Biciclo [3.1.1]heptan-3-ona, 2,6,6-trimetil-, (1R, 2R, 5S)-rel - (Iso-pinocanfona)	1187				4,1596
62	Fenol, 4-(metoximetil) - ^① (2-metoxi-p-Cresol)	1191 *		0,0095		
63	3-Ciclohexeno-1-metanol, $\alpha,\alpha,4$ -trimetil - (α -Terpineol)	1200	0,0591	0,0139		0,3560
64	Biciclo [3.1.1]hept-2-eno-2-carboxaldehído, 6,6-dimetil - (Mirtenal)	1205	0,4781			0,1290
65	Biciclo [3.1.1]hept-3-en-2-ona, 4,6,6-trimetil-, (1R, 5R) - (Verbenona)	1217	0,6711			0,0381



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus Salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
66	Benceno, 2-metoxi-4-metil-1-(1-metiletil)- (Eter metil-timol)	1223				0,1156
67	1-Ciclohexeno-1-carboxaldehído ,2,6,6-trimetil- (β-Ciclocitral)	1227	0,0728		0,0113	
68	Benceno, 2-metoxi-1-metil-4-(1-metiletil) - (Eter metil-carvacrol)	1229				0,0734
69	Biciclo [2.2.1]heptan-2-ol, 1,7,7-trimetil-, 2-formiato, (1R, 2R, 4R)-rel - ^① (Isobornil formato)	1238		0,0054		
70	Ciclohexanona, 5-metil-2 -(1-metiletiliden) - ^① (Pulegona)	1246			0,0049	
71	Fenol, 4-(2-propen-1-il)- ^① (Cavicol)	1249		0,0343		
72	2-Ciclohexen 1-ona, 2-metil-5- (1-metiletenil) - (Carvona)	1250	0,0636			
73	Benzaldehído, 4-(1-metiletil) - (Cuminaldehido)	1251				0,1240
74	Fenol, 4-etil-2-metoxi - (4-etil Guayacol)	1277		0,0090		
75	Biciclo [2.2.1] heptan-2-ol, 1,7,7-trimetil-, 2-etilo, (1R, 2S, 4R)-rel - (Acetato de bornilo)	1291	6,3269			1,6941
76	Biciclo [3.1.1] heptan-3-ol, 6,6-dimetil-2-metileno-, 3-etilo,(1R, 3S,5R) - ^① (Acetato pinocarvyl (cis-))	1299 *	0,0337			



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
77	Biciclo [3.1.1]hept-2-eno-2-metanol, 6,6-dimetil-,2-etilo, (1S, 5R) - (Acetato mirtenil)	1326	0,1079			
78	Biciclo [4.1.0] heptano, 3-etnil-3,7,7-trimetil-2-(1-metiletenil)-, (1R, 2R, 3R, 6R)- ^① (Bicicloelemeno)	1335		0,0116		
79	Ciclohexeno, 4-etnil-4-metil-3-(1-metiletenil)-1-(1-metiletil)-, (3R, 4R) (δ-Elemeno)	1342		0,0115		
80	2-Propinoico ácido, 3-, fenil éster etílico ^① (Acetato 3-fenilpropanoato)	1348 *	0,0830			
81	Fenol, 2-metoxi-4-(2-propen-1-il)- (Eugenol)	1355		0,0472		
82	1H-Ciclopenta [1,3] ciclopropa [1,2] benceno, 3a, 3b, 4, 5, 6, 7- hexahidro- 3, 7-dimetil-4-(1-metiletil)-, (3aS, 3br, 4S, 7R, 7aR) (α-Cubebena)	1355	0,0258			0,0281
83	Triciclo [4.4.0.0(2,7)] dec-3-eno, 1,3-dimetil-8-(1-metiletil)-, (1S, 2R, 6R, 7R, 8S)- (α-Ylangeno)	1381		0,0381		
84	1,2,4-Metheno-1H-indeno, octahidro-1,7a-dimetil-5-(1-metiletil)-, (1S, 2S, 3aR, 4R, 5S, 7aS, 8R) - (Ciclozativeno)	1386	0,1760			1,0887
85	Triciclo [4.4.0.0(2,7)]dec-3-ene, 1,3-dimethyl-8-(1-methylethyl)-, (1R,2S,6S,7S,8S)-9 (α-Copaeno)	1388		0,2244		
86	Benceno, 1,2-dimetoxi-4-(2-propen-1-il) - (Metil-eugenol)	1395				0,0667
87	Ciclobuta [1,02:03,4] dicyclopentene, decahidro-3a-metil-6-metileno-1-(1-metiletil)-, (1S, 3aS, 3br, 6aS, 6br) - (β-Bourboneno)	1398		0,8777		



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus Salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
88	1H-Cicloprop[e]azuleno, 1bis, 2, 3, 4, 4a, 5, 6, 7b-octahidro-1, 1, 4, 7-tetrametil-, (1aR, 4R, 4aR, 7Bs)- (α -Gurjuneneno)	1423	0,0244	0,0553		
89	1,4,8-Cycloundecatriene, 2,6,6,9-tetrametil-, (1E, 4E, 8E)- (E-Cariofileno)	1437	0,0726	1,2739	0,0058	0,6372
90	Triciclo [4.4.0.0(2,7)] decano,1-metil-3-metileno-8-(1-metiletil)-, (1R, 2S, 6S, 7S, 8S) -rel - (β -Copaeno)	1445 *		0,0841		
91	1,6,10-Dodecatrieno,7,11-dimetil-3-metileno-, (6Z)- ((Z)- β -Farneseno)	1447				0,0462
92	Ciclohepteno, 1,7-dimetil-7-(4-metil-3-penten-1-il)- (α -Santaleno)	1455		0,0070		
93	Azuleno, (1R, 3aR, 4R, 7R)- (γ -Gurjuneneno)	1457	0,0760			
94	1H-Cicloprop[e]azuleno, decahidro-1,1,7-trimetil-4-metilen-, (1aR, 4aS, 7R, 7aR, 7BS)- (Alloaromadendreno)	1478		0,0567		
95	Biciclo [7.2.0]undec-4-eno, 4,11,11-trimetil-8-metileno-, (1R, 4E, 9R) - (epi- β -Cariofileno)	1479	0,2033			0,0723
96	Benceno, 1-[(1R)-1,5-dimetil-4-hexenilo]-4-metil - (ar-Curcumeno)	1482		0,2123		
97	Naftaleno, (1R, 4aR, 8aS)-rel - (γ -Muuroleno)	1486				0,1079



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
98	Naftaleno, 1,2,4a,5,6,8a-hexahidro-4,7-dimetil-1-(1-metiletil) - ^① (α-Amorfenoleno)	1489		0,1073		
99	1,3-Ciclohexadieno, 5-[(1S)-1,5-dimetil-4-hexen-1-il]-2-metil-, (5R) - (α-Zingibereno)	1494		0,1120		
100	1H-Cicloprop [e]azuleno,1bis, 2, 3, 5, 6, 7, 7a, 7b-octahidro-1,1,4,7-tetrametil-, (1aR, 7R, 7aS, 7bR) - (Viridiflorenoleno)	1506	0,3451			
101	Naftaleno, 1,2,4a,5,6,8a-hexahidro-4,7-dimetil-1-(1-metiletil)-, (1R, 4aR, 8aS)-rel - ^① (α-Muuroloeno)	1507		0,1241		0,1746
102	Naftaleno, (2R, 4aR, 8aR) - (α-Selineno)	1512		0,0273		
103	Naftaleno, 1,2,3,5,6,8a-hexahidro-4,7-dimetil-1-(1-metiletil)-, (1S, 8aS) - ^① (δ-Amorfenoleno)	1526 *		0,1302		0,2166
104	Naftaleno, 1,2,3,5,6,8a-hexahidro-4,7-dimetil-1-(1-metiletil)-, (1S, 8aR) - (δ-Cadineno)	1527	0,0322			
105	Naftaleno, 1,2,3,4-tetrahidro-1,6-dimetil-4-(1-metiletil)-, (1S, 4S) - (Calameneno)	1533		0,0369		0,4273
106	Naftaleno, 1,2,3,4,5,6-hexahidro-4,7-dimetil-1-(1-metiletil)-, (1S, 4R) - ^① (Cadinoleno-1,4-dieno (trans-))	1544 *		0,0118		0,0223
107	Naftaleno, 1,2,4a,5,6,8a-hexahidro-4,7-dimetil-1-(1-metiletil)-, (1S, 4aR, 8aR) - (α-Cadineno)	1548		0,0087		



#	Compuesto	IK	<i>Cistus ladanifer</i> (% área)	<i>Cistus albidus</i> (% área)	<i>Cistus salviifolius</i> (% área)	<i>Cistus clusii</i> (% área)
108	Naftaleno, 1,2-dihidro-4,7-dimetil-1-(1-metiletil)-, (1S)- (α -Calacoreno)	1554		0,0235		0,0541
109	Naftaleno, (4aR, 8aR)- (Selina-3,7(11)-dieno)	1559		0,0093		
110	Espiro [4.5]dec-8-ES-7-uno, 1,8-dimetil-4-(1-metiletil)-, (1S, 4S, 5S) - ^① (Acorenona)	1569 *		0,0122		
111	5-Oxatriciclo[8.2.0.0(4,6)] dodecano, 4,12,12-trimetil-9-metilen-, (1R, 4R, 6R, 10S) - (Oxido de cariofileno)	1604				0,0971
112	1H-Cicloprop[e]azulen-4-ol , decahidro-1,1,4,7-tetrametil-, (1aR, 4S, 4aS, 7R, 7aS, 7BS) - (Viridiflorol)	1616	0,2301			

El % área relativo a la suma de las áreas de todos los compuestos.

IK, índice de retención de Kovats.

^① Compuestos identificados por primera vez en *Cistus*.

* Compuestos identificados sólo por el espectro de masas (IK no coincide).



Entre las numerosas ventajas proporcionadas por equipo automatizado de SPME-GC-MS, destaca el poder prescindir totalmente de la utilización de disolventes dado que la desorción térmica se produce directamente desde la fibra al equipo de GC. El sistema de agitación automatizado también constituye una ventaja frente a la posibilidad de contaminación utilizando agitadores magnéticos externos. La fibra utilizada proporcionó buenos resultados en la extracción de volátiles. Estudios previos han descrito el uso de este tipo de fibras de doble recubrimiento DVB/PDMS como un método simple con el que se obtienen altos rendimientos de extracción, ofreciendo una alta selectividad y estabilidad térmica²⁰². Se ha demostrado que esta fibra es adecuada para compuestos volátiles dentro de un rango de masas de m/z 40-270⁷³.

El modo de extracción *headspace* (HS-SPME) utilizado ha demostrado ser especialmente útil para la extracción de compuestos volátiles, evitando interferencias de alto peso molecular. Esta característica también puede considerarse una desventaja en el sentido de que imposibilita la extracción y por tanto, la identificación de moléculas más pesadas que podrían resultar de interés.

De acuerdo a las clases químicas presentes en las muestras analizadas, los monoterpenos y sesquiterpenos dominaron el perfil de volátiles. Entre los monoterpenos, se distinguen diversos tipos en función de los grupos funcionales que poseen en su estructura. La mayoría son hidrocarburos monoterpénicos, aunque también se han identificado cetonas, ésteres y aldehídos monoterpénicos. Además, se han encontrado otras clases de compuestos de naturaleza no isoprenoide como alcoholes, cetonas y aldehídos. Sin embargo, no se identificaron compuestos de naturaleza diterpénica en este

²⁰² Elmore, J.S., Mottram, D.S. & Hierro, E., 2001. Two-fibre solid-phase microextraction combined with gas chromatography-mass spectrometry for the analysis of volatile aroma compounds in cooked pork. *Journal of Chromatography A*, 905(1-2), pp.233–240.



estudio, lo que concuerda con los resultados descritos por otros autores utilizando el mismo método de extracción²⁰³. Los diterpenos, con cuatro unidades de isopreno en su estructura, requieren una mayor energía para volatilizarse. Para su recuperación total, sería necesario obtener el aceite esencial de la planta mediante un largo proceso de destilación, o bien obtener los extractos hexénicos. En ambos casos, el análisis se realizaría mediante inyección directa en el equipo cromatográfico, obviando la etapa de SPME. Según la bibliografía revisada, se han encontrado diterpenos tipo labdano en aceites esenciales de *C. salviifolius* obtenidos mediante hidrodestilación²⁰⁴. También se han descrito una gran cantidad de diterpenos en un estudio realizado mediante extracción con hexano de las especies de *Cistus libanotis*, *C. villosus* y *C. monspeliensis*²⁰⁵.

En cuanto a la comparación cualitativa y cuantitativa entre los distintos extractos, cabe destacar que el extracto con el mayor número de compuestos identificados fue *C. ladanifer*. Como puede observarse tanto en los perfiles como en la tabla de compuestos identificados, el extracto más rico en sesquiterpenos fue *C. albidus* y, por el contrario, en *C. salviifolius* sólo se identificó un compuesto sesquiterpénico, el E-cariofileno. A partir de este compuesto, identificado a 39,9 min, no se identificó ningún compuesto en *C. salviifolius*.

En total, se identificaron 112 compuestos volátiles, de los cuales 28 son identificados por primera vez en *Cistus*. Entre estos compuestos nuevos, se encontraron principalmente compuestos de naturaleza terpénica, aunque

²⁰³ Maccioni, S. *et al.*, 2007. *In vivo* volatiles emission and essential oils from different organs and pollen of *Cistus albidus* from Caprione (Eastern Liguria, Italy). *Flavour and Fragrance Journal*, 22, pp.61–65.

²⁰⁴ Demetzos, C., Angelopoulou, D. & Perdetzoglou, D., 2002. A comparative study of the essential oils of *Cistus salviifolius* in several populations of Crete (Greece). *Biochemical Systematics and Ecology*, 30(7), pp.651–665.

²⁰⁵ Ben Jemia, M. *et al.*, 2013. Antiproliferative activity of hexane extract from Tunisian *Cistus libanotis*, *Cistus monspeliensis* and *Cistus villosus*. *Chemistry Central Journal*, 7(1), pp.47–53.



también algunos compuestos volátiles no terpénicos al principio del cromatograma. Entre estos últimos, se distinguen cetonas como la 3-pentanona (en *C. clusii*), alcoholes como el 3-metil-1-butanol (*C. albidus*) y aldehídos como el 3-metil-2-butenal (*C. ladanifer*).

Entre los terpenos identificados por primera vez en esta planta, destaca el p-cimeno, detectado en todas las especies analizadas a excepción de *C. salviifolius* y presentando un área mayor en *C. clusii* (0.33%). También se describen por primera vez los compuestos monoterpénicos trans-sabinol y pulegona en *C. clusii* y *C. salviifolius*, respectivamente. Aunque presentando un valor de área relativamente bajo, cabe destacar otros compuestos nuevos como el veratrol (pirocatecol dimetiléter) en *C. clusii* (0.11%), y el cavicol, un monoterpene fenólico detectado únicamente en *C. albidus* (0.03%). Al final del cromatograma, se identificaron compuestos nuevos de naturaleza sesquiterpénica como el bicicloelemeno (0.01%), el δ -amorfenol (0.13%), el α -amorfenol (0.11%) y la acorenona (0.01%), todos ellos presentes en *C. albidus*. El compuesto trans-cadina-1,4-dieno también es un sesquiterpene descrito por primera vez en *Cistus*, identificado concretamente en *C. albidus* (0.01%) y *C. clusii* (0.02%).

El resto de compuestos han sido descritos por otros autores. La mayoría de los compuestos mayoritarios identificados, se encuentran previamente descritos en *C. ladanifer* y otras especies del género *Cistus*. Sin embargo, se ha de matizar que muchos de estos compuestos no se refieren a las mismas especies que conforman el presente estudio. El β -pineno, el alcanfor, el ciclozatenol y el E-cariofileno se han descrito en las especies *C. ladanifer*, *C. salviifolius* y *C. albidus*, además de otras muchas especies como *C. incanus*, *C. parviflorus*, *C.*



monspeliensis, *C. creticus*, *C. villosus*, etc²⁰³⁻²¹⁰. A este respecto, se ha de destacar la identificación de β -pineno, alcanfor y E-cariofileno por primera vez en *C. clusii*, apreciándose además un alto valor de área para dichos compuestos. Como ejemplos de compuestos descritos en *C. ladanifer* y *C. salviifolius*, se encuentran el canfeno, la pinocarvona y el acetato de bornilo^{207,211}. Estos compuestos se han identificado en la muestra de *C. clusii* analizada, siendo por tanto la primera vez que se describen en esta especie. Asimismo, compuestos como el α -pineno, el limoneno y el p-cimeno se encuentran descritos en *C. ladanifer* y *C. albidus*, sin embargo, se han identificado por primera vez en las muestras de este trabajo, concretamente el limoneno en *C. clusii*, el p-cimeno en *C. salviifolius*, y el α -pineno en *C. clusii* y *C. salviifolius*^{203,208,209}. La iso-pinocanfona es una cetona monoterpénica encontrada exclusivamente en *C. clusii* a una cantidad alta (4.16%). Constituye otro ejemplo de compuesto previamente descrito en *C. ladanifer*, pero no en *C. clusii*. Además de haberse identificado en *C. ladanifer*, la iso-pinocanfona se ha descrito únicamente en *C. monspeliensis* y *C. laurifolius*^{206,212}.

Los constituyentes volátiles mayoritarios de las muestras analizadas fueron el canfeno, el β -pineno y el α -pineno (11.74%, 10.49% y 7.21% en *C. clusii*,

²⁰⁶ Ögütveren, M. & Tetik, S.S., 2004. Composition of the essential oil of *Cistus parviflorus* L. from Turkey. *Journal of Essential Oil Research*, 16(2), pp.115–116.

²⁰⁷ Loizzo, M.R. et al., 2013. Chemistry and functional properties in prevention of neurodegenerative disorders of five *Cistus* species essential oils. *Food and chemical toxicology*, 59, pp.586–594.

²⁰⁸ Viuda-Martos, M. et al., 2011. Identification of flavonoid content and chemical composition of the essential oils of Moroccan herbs: myrtle (*Myrtus communis* L.), rockrose (*Cistus ladanifer* L.) and Montpellier cistus (*Cistus monspeliensis* L.). *Journal of Essential Oil Research*, 23(2), pp.1–9.

²⁰⁹ Llusà, J. et al., 2009. Annual and seasonal changes in foliar terpene content and emission rates in *Cistus albidus* L. submitted to soil drought in Prades forest (Catalonia, NE Spain). *Acta Physiologiae Plantarum*, 32(2), pp.387–394.

²¹⁰ Rincón, J., De Lucas, A. & Gracia, I., 2000. Isolation of rock rose essential oil using supercritical CO₂ extraction. *Separation Science and Technology*, 35(16), pp.2745–2763.

²¹¹ Verdeguer, M., Blázquez, M.A. & Boira, H., 2012. Chemical composition and herbicidal activity of the essential oil from a *Cistus ladanifer* L. population from Spain. *Natural Product Research*, 26(17), pp.1602–1609.

²¹² Oller-López, J.L. et al., 2005. Composition of the essential oils of *Cistus ladaniferus* and *C. monspeliensis* from Morocco. *Journal of Essential Oil Research*, 17(5), pp.553–555.



respectivamente). Estos hidrocarburos monoterpénicos se encuentran también en *C. ladanifer*, aunque presentando un área menor. El limoneno también destaca en la composición de *C. clusii* (5.50%), habiéndose identificado en *C. ladanifer* y *C. albidus*. El acetato de bornilo es un éster monoterpénico que se ha identificado tanto en *C. clusii* como en *C. ladanifer*, aunque en esta última especie con un área mayor (6.33%).

Además de los compuestos mayoritarios anteriormente citados, se han identificado muchos compuestos con área relativamente alta tales como el α -tujeno en *C. ladanifer* (0.86%). El alcanfor es una cetona de cuya identificación destaca por ser el único monoterpeno identificado en las cuatro especies de *Cistus*, con mayor porcentaje de área en *C. ladanifer* (0.37%). El p-cimeno se identificó en tres de la especies analizadas, en *C. albidus*, *C. salviifolius* y *C. ladanifer*, siendo más abundante en esta última (1.83%). Del mismo modo, el γ -terpineno se identificó en *C. ladanifer*, *C. salviifolius* y *C. clusii* (0.54%, 0.01% y 2.31%, respectivamente). La pinocarvona es una cetona monoterpénica que se identificó en *C. ladanifer* (1.26 %) y en *C. clusii* (0.17 %).

En cuanto a la identificación de sesquiterpenos, son todos de área baja o relativamente alta y aparecen a partir del minuto 36 del cromatograma. Cabe destacar la identificación de ciclozativeno en *C. ladanifer* (0.18 %) y *C. clusii* (1.09%) y el β -bourboneno, que sólo se encontró en *C. albidus* (0.88%). El E-cariofileno es el único sesquiterpeno común a las cuatro especies de *Cistus*, siendo más abundante en *C. albidus* y *C. clusii*.

Por último, otros compuestos de naturaleza no terpénica que también se identificaron fueron, por ejemplo, el 3Z-hexenol (sólo en *C. albidus* a 0.44%), la trans-3-nonen-2-ona (sólo en *C. ladanifer* a 0.24%) y el o-guayaacol (sólo en *C.*



albidus a 1.37%). Este último compuesto se encuentra descrito en bibliografía correspondiente a *C. salviifolius*²¹³.

2.2. Estudio de la fracción polar de especies del género *Cistus*.

Para el estudio de la capacidad antimicrobiana de especies de *Cistus*, se optó por la fracción más polar. Se estudió la capacidad antimicrobiana y la composición polifenólica de extractos de cuatro especies distintas del género *Cistus* (*C. ladanifer*, *C. albidus*, *C. salviifolius* y *C. clusii*), siendo la primera vez que se describe la capacidad antibacteriana de extractos acuosos de *C. salviifolius*. En este estudio, se presentan primeramente los resultados de la valoración antimicrobiana y posteriormente la caracterización de los extractos polifenólicos de *Cistus* incluidos en el estudio. Para completar este apartado de caracterización, se realizó un estudio comparativo entre los resultados de la caracterización convencional y la caracterización utilizando los datos procesados con el método propuesto de procesamiento de datos de MS. Por último, se presentan los resultados del estudio de correlación entre ambos datos, la capacidad antimicrobiana y el contenido polifenólico, con el fin de encontrar el/los compuesto/s potencialmente responsable/s de dicha actividad biológica.

En la preparación de extractos polifenólicos, las especies estudiadas fueron las mismas que las del estudio anterior de la fracción apolar. Se utilizaron distintos sistemas de extracción (maceración acuosa y maceración hidroalcohólica) y de secado (liofilización y spray-drying), así como también se incluyeron procesos de fraccionamiento con el fin de obtener extractos más ricos en polifenoles. La combinación entre las diferentes especies de *Cistus* incluidas y los diferentes

²¹³ Tinoco, P., Almendros, G. & Gonza, F.J., 2002. Impact of the vegetation on the lignin pyrolytic signature of soil humic acids from Mediterranean soils. *Journal of Analytical and Applied Pyrolysis*, 64, pp.407–420.



procesos aplicados para la obtención de los extractos supone una colección de muestras más amplia que la del estudio de los compuestos volátiles.

2.2.1. Capacidad antimicrobiana de los extractos polifenólicos de *Cistus*.

Se evaluó la capacidad antimicrobiana de los extractos polifenólicos de *Cistus* mediante la utilización de un modelo de bacteria Gram positiva (*S. aureus*) y un modelo de bacteria Gram negativa (*E. coli*). En la Tabla 29 se presentan los resultados obtenidos en el ensayo de valoración de la capacidad antimicrobiana para todos los extractos. Se presentan por separado, según el modelo de bacteria utilizado y se expresan como concentración mínima inhibitoria 50 (MIC₅₀), es decir, la concentración mínima de extracto requerida para inhibir el 50% del crecimiento bacteriano. Por combinación entre las cuatro especies incluidas, los dos métodos de extracción utilizados y los dos métodos de secado llevados a cabo, se obtiene un total de 12 extractos, tres de cada especie de *Cistus*. Los códigos de los mismos aparecen también en la tabla.



Tabla 29. Capacidad antimicrobiana de los extractos de *Cistus* frente a *E. coli* (modelo de bacteria Gram negativa) y *S. aureus* (modelo de bacteria Gram positiva), expresada como concentración mínima inhibitoria 50 (MIC₅₀) en µg/mL.

Especies de <i>Cistus</i>	Extracción/Secado	Extracto	<i>E. coli</i> MIC ₅₀	<i>S. aureus</i> MIC ₅₀
<i>C. ladanifer</i>	H ₂ O/ FD	L1	612 ± 94	569 ± 112
	H ₂ O/ SDY	L2	248 ± 47	151 ± 19
	H ₂ O-EtOH/ FD	L3	113 ± 23	144 ± 16
<i>C. albidus</i>	H ₂ O/ FD	A1	336 ± 106	273 ± 142
	H ₂ O/ SDY	A2	504 ± 129	60 ± 8
	H ₂ O-EtOH/ FD	A3	233 ± 45	292 ± 24
<i>C. clusii</i>	H ₂ O/ FD	C1	1064 ± 350	392 ± 56
	H ₂ O/ SDY	C2	501 ± 173	91 ± 23
	H ₂ O-EtOH/ FD	C3	116 ± 34	304 ± 35
<i>C. salviifolius</i>	H ₂ O/ FD	S1	221 ± 40	52 ± 18
	H ₂ O/ SDY ¹	S2	289 ± 96	50 ± 2
	H ₂ O-EtOH/ FD	S3	245 ± 91	45 ± 11
	H ₂ O/ SDY ¹ fracción retenida	S2a	14 ± 2	11 ± 1
	H ₂ O/ SDY ¹ fracción lavada	S2b	195 ± 24	77 ± 89
	H ₂ O/ SDY ²	S4	106 ± 16	87 ± 13

H₂O, maceración acuosa.

H₂O-EtOH, maceración hidroalcohólica.

FD, freeze-drying o liofilización.

SDY, spray-drying.

S2a y S2b proceden del fraccionamiento de S2.

¹Material vegetal recolectado en Noviembre.

²Material vegetal recolectado en Agosto-Septiembre.

Se seleccionó el procedimiento de extracción con agua seguido de secado spray-drying como un procedimiento de extracción rentable y ecológico, además de ser más adecuado a escala industrial que la liofilización. Concretamente, con el fin de obtener un extracto polifenólico purificado, se seleccionó el extracto H₂O/SDY de *C. salviifolius* (S2) debido a la fuerte actividad inhibitoria que presentó contra *S. aureus*. Se fraccionó utilizando una columna cromatográfica



afín a la fracción polifenólica, conteniendo una resina Amberlite. De este modo, se incluyeron al estudio dos extractos adicionales: la fracción polifenólica retenida en la columna y eluída con etanol (S2a), y la fracción correspondiente al lavado de la columna con agua (S2b).

Además de los extractos resultantes del fraccionamiento del extracto bioactivo S2, se incluyó un último extracto al estudio de caracterización (denominado S4 puesto que también procede de *C. salvifolius*). La estacionalidad de las plantas de *Cistus* supone diferencias sustanciales en cuanto a la composición fitoquímica de las mismas, de modo que para obtener el extracto S4 se llevó a cabo el mismo procedimiento seguido con S2 (H₂O/SDY), pero variando el tiempo de recolección. El material vegetal para S4 se adquirió de la misma plantación que S2 pero en vez de Noviembre, se recolectó en Agosto-Septiembre, después de la estación floreciente²¹⁴⁻²¹⁶.

Se evaluó la capacidad antibacteriana para los extractos incorporados al estudio en último lugar, S2a, S2b y S4. Dichos datos se detallan en la parte final de la Tabla 29. La extracción hidroalcohólica y el método de secado spray-drying para *C. ladanifer* (L3) mostró una capacidad antibacteriana más fuerte contra ambos modelos bacterianos. El extracto acuoso de esta especie secado mediante spray-drying (L2) fue mucho más potente contra *S. aureus* que contra *E. coli*, lo cual concuerda con la información publicada en artículos previos⁵¹.

Los extractos de *C. albidus* y *C. clusii* presentaron perfiles similares, en el sentido de que la actividad antibacteriana fue mayor contra *E. coli* para los extractos hidroalcohólicos (A3, C3) y más fuerte contra *S. aureus* para el caso de los

²¹⁴ Chaves, N., Escudero, J.C. & Rrez-merinot, C.G., 1997. Quantitative variation of flavonoids among individuals of a *Cistus ladanifer* population. *Biochemical Systematics and Ecology*, 25(5), pp.429–435.

²¹⁵ Chaves, N., Sosa, T. & Escudero, J.C., 2001. Plant growth inhibiting flavonoids in exudate of *Cistus ladanifer* and in associated soils. *Journal of Chemical Ecology*, 27(3), pp.623–631.

²¹⁶ Sosa, T. et al., 2005. Interpopulational variation in the flavonoid composition of *Cistus ladanifer* L. exudate. *Biochemical Systematics and Ecology*, 33(4), pp.353–364.



extractos acuosos secados con SDY (A2, C2). Cabe destacar que el extracto A2 fue especialmente potente contra *S. aureus*, mostrando valores MIC₅₀ tan bajos como 60 µg/mL.

Los extractos de *C. salviifolius* mostraron la actividad más fuerte contra *S. aureus* de todas las muestras analizadas, con valores MIC₅₀ muy bajos, entre 45 y 52 µg/mL, dependiendo de los procesos de extracción y secado utilizados.

Como comparación entre los procedimientos utilizados, la extracción hidroalcohólica (extractos H₂O-EtOH/FD) destacó por una mayor inhibición del crecimiento de *E. coli*, excepto en el caso de *C. salviifolius*, cuyos valores MIC₅₀ fueron menores en los extractos acuosos (S1 y S2).

Con respecto a la inhibición del crecimiento de *S. aureus*, los mejores resultados se obtuvieron con la extracción acuosa seguida de secado spray-drying, especialmente en *C. salviifolius*, *C. clusii* y *C. albidus*, lo cual indica que muchos compuestos bioactivos fueron extraídos y preservados bajo estas condiciones polares.

Como comparación entre los métodos de extracción/secado utilizados, es interesante destacar que los extractos acuosos liofilizados mostraron valores de actividad antimicrobiana bajos, especialmente contra *S. aureus*, en comparación con los extractos acuosos secados con spray-drying, no siendo especialmente notable este efecto para los extractos de *C. clusii* y *C. ladanifer*. Este resultado reveló que algunos compuestos polares que contribuyen a la inhibición de *S. aureus* se degradan en el proceso. Las explicaciones más convincentes pueden ser o bien, la degradación de algunos compuestos bioactivos en el paso de concentración en el que la temperatura es de 50-60°C, o bien, la eliminación de



compuestos volátiles tales como los terpenos, de los cuales es bien conocida la actividad antimicrobiana^{199,217}.

En general, los resultados obtenidos mostraron que sin duda la actividad biológica se vio fuertemente afectada por los procedimientos de extracción y secado utilizados. Sin embargo, el quimiotipo de las especies de *Cistus* fue otra variable a considerar, ya que los extractos derivados de *C. salviifolius* fueron los más bioactivos contra *S. aureus*, independientemente del procedimiento de extracción/secado empleado.

El extracto S2 mostró importantes efectos bacteriostáticos contra *S. aureus*, en comparación con el resto de extractos de *Cistus* derivados de especies españolas. La actividad antimicrobiana de este extracto se incrementó considerablemente, tanto contra bacterias Gram positivas como Gram negativas, cuando se purificó el extracto. La fracción S2a presentó una fuerte actividad antimicrobiana contra ambas cepas bacterianas, especialmente contra *E. coli*. La fracción S2b mostró una actividad mayor contra *E. coli* que el extracto S2 y aun mantenía una actividad inhibitoria significativa contra *S. aureus*. Las curvas dosis-respuesta para el extracto S2 y la fracción purificada S2a (Figura 92) mostraron que en el caso de *E. coli*, el mecanismo de acción es más selectivo, al obtenerse una pendiente de la curva más abrupta para S2a con respecto al S2.

²¹⁷ Greche, H. *et al.*, 2009. The volatiles of the leaf oil of *Cistus ladanifer* L. var. *albiflorus* and Labdanum extracts of Moroccan origin and their antimicrobial activities. *Journal of Essential Oil Research*, 21(2), pp.166–173.



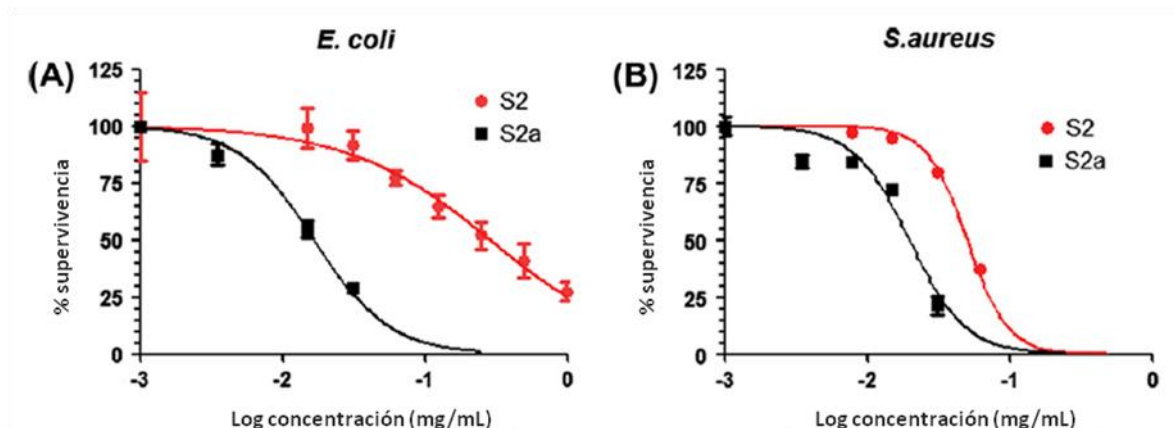


Figura 92. Gráficas de supervivencia de *E. coli* (A) y *S. aureus* (B) en presencia de concentraciones crecientes de los extractos de *C. salviifolius* S2 (extracto acuoso secado con spray-drying) y S2a (fracción fenólica derivada de S2).

El extracto S4 mostró una capacidad antibacteriana contra *E. coli* mayor que S2, casi comparable a la obtenida con el extracto purificado S2a. Esto puede deberse a un aumento en la síntesis de polifenoles como respuesta a una mayor exposición a la radiación UV durante el verano. El extracto purificado S2a mostró incrementos en su capacidad para inhibir el crecimiento de *E. coli* y *S. aureus* de 20 veces y 4.5 veces más que S2, respectivamente.

Los valores MIC₅₀ obtenidos en este estudio para los extractos acuosos de *C. salviifolius* fueron más bajos que los citados para aceites esenciales de *C. salviifolius*¹⁹⁹ o aquellos descritos para extractos acuosos derivados de especies de *Cistus* del mismo subgénero (*Leucocistus*)⁵¹. Los valores MIC₅₀ para el extracto acuoso de *C. salviifolius* (50.0 µg/mL) y su fracción polifenólica purificada (11.0 µg/mL) revelaron una mejorada actividad bacteriostática contra *S. aureus*, comparada con el antibiótico utilizado (100.0 µg/mL de neomicina) para ambos extractos^{51,218}.

²¹⁸ Laporta, O. et al., 2007. Isolation, characterization and antioxidant capacity assessment of the bioactive compounds derived from *Hypoxis rooperi* corm extract (African potato). *Food Chemistry*, 101(4), pp.1425–1437.



2.2.2. Caracterización del perfil polifenólico de los extractos de *Cistus*.

Se analizaron los extractos de *Cistus* mediante RRLC-ESI-TOF-MS y se obtuvieron los perfiles de los mismos. Tres de los extractos, aquellos derivados de *C. salviifolius* (S2, S2a y S4), se seleccionaron por mostrar una alta capacidad antimicrobiana en el estudio anterior de bioactividad y con ellos se procedió al estudio de caracterización pormenorizada.

Los perfiles de estos extractos de *C. salviifolius* se muestran de manera separada en la Figura 93. Los resultados de esta caracterización se presentan en la Tabla 30.



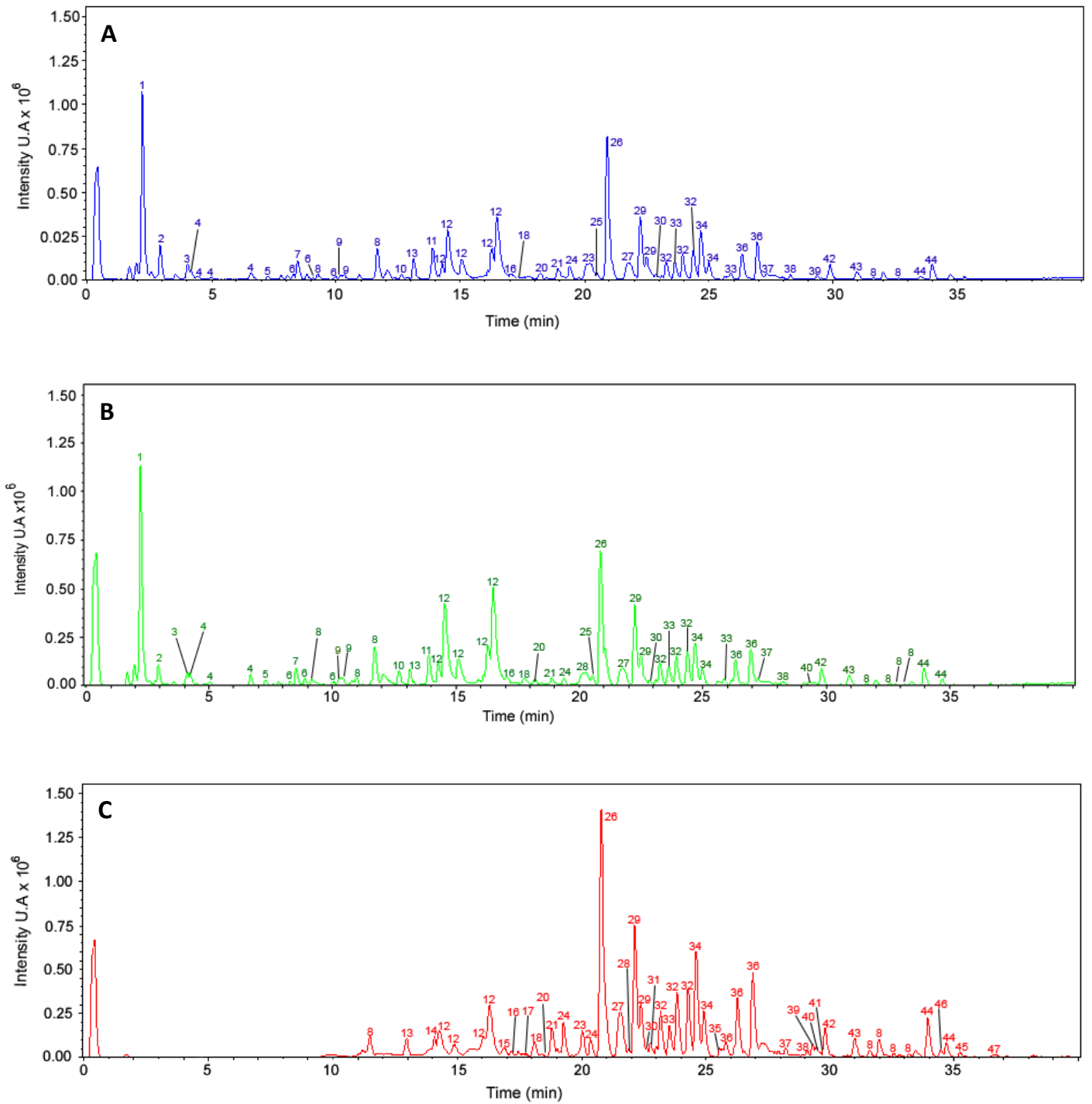


Figura 93. Cromatogramas de pico base de los extractos de *C. salviifolius*. (A) S2, (B) S4, (C).

S2a.



Tabla 30. Compuestos identificados en los extractos de *C. salviifolius* analizados mediante RRLC-ESI-TOF-MS.

#	[M-H] ⁻	t _R (min)	Fórmula [M-H] ⁻	Error (ppm)	mSigma	ISCID	Compuesto propuesto	Área		
								S2	S2a	S4
1	191,0552	2,26	C 7 H 11 O 6	11,2	4,9	-	Ácido quínico	8825814		9453396
2	173,0451	2,98	C 7 H 9 O 5	10,8	3,4	-	Ácido shikímico	1510173		804404
3	337,0750	4,08	C 12 H 17 O 11	14,2	5,0	-	Ácido ascórbico glicósido ^①	738455		544406
4	481,0613	4,21	C 20 H 17 O 14	8,2	4,0	-	Hexahidroxidifenoil-D-glucosa I (isómero)	420910		566283
5	481,0607	4,77	C 20 H 17 O 14	13,7	8,3	-	Hexahidroxidifenoil-D-glucosa II (isómero)	53670		106070
6	481,0604	5,00	C 20 H 17 O 14	11,7	29,3	-	Hexahidroxidifenoil-D-glucosa III (isómero)	150028		208650
7	481,0638	6,62	C 20 H 17 O 14	12,1	8,2	-	Hexahidroxidifenoil-D-glucosa IV (isómero)	385317		465430
8	271,0801	7,29	C 12 H 15 O 7	6,1	4,5	-	Arbutina ^①	129424		234484
9	331,0654	8,09	C 13 H 15 O 10	5,4	28,3	-	Glucogalina I (isómero)	167638		25008
10	169,0137	8,50	C 7 H 5 O 5	12,7	5,7	-	Ácido gálico	952231		751483
11	609,1200	9,18	C 30 H 25 O 14	3,7	8,5	-	(-)-(Epi)galocatequina-(epi)galocatequina dímero I (isómero)	75151	322421	220759
12	609,1228	10,82	C 30 H 25 O 14	4,1	18,0	-	(-)-(Epi)galocatequina-(epi)galocatequina dímero II (isómero)		457609	209865
13	609,1223	11,74	C 30 H 25 O 14	1,3	4,3	-	(-)-(Epi)galocatequina-(epi)galocatequina dímero III (isómero)	1672211	1187673	1942534
14	331,0635	8,87	C 13 H 15 O 10	2,4	7,0	-	Glucogalina II (isómero)	231642		307276
15	331,0648	10,02	C 13 H 15 O 10	0,9	8,7	-	Glucogalina III (isómero)	145472		195056
16	781,0483	10,23	C 34 H 21 O 22	12,3	18,6	-	Punicalina I (isómero)	151965		257665
17	781,0494	10,35	C 34 H 21 O 22	10,3	17,9	-	Punicalina II (isómero)	283800		390862
18	305,0629	12,66	C 15 H 13 O 7	6,4	12,5	-	Epigalocatequina	180254		742976
19	285,0586	13,91	C 12 H 13 O 8	1,1	0,8	-	Uralenneósido	1686905		1465160
20	1083,0419	14,52	C 48 H 27 O 30	2,3	7,5	541,0229	Punicalagina I (isómero)	3861955	2119931	5728138
21	1251	15,07	-	-	-	603,1903	Punicalagina galato	1549911	1018042	1731421
22	761,1343	14,11	C 37 H 29 O 18	1,9	4,6	-	Prodelfinidina B2, 3'-O-galato	840094	1378171	993475



23	1083,0404	16,49	C 48 H 27 O 30	2,5	21,5	541,0217	Punicalagina II (isómero)	4769833	4532601	6861677
24	1085,0780	16,89	C 48 H 29 O 30	0,1	35,0	542,0345	Cornusiina B	250560	652657	381170
25	289,0684	17,32	C 15 H 13 O 6	3,8	3,6	-	(-)-(Epi)catequina	22876	229041	17170
26	329,0880	17,67	C 14 H 17 O 9	-0,6	20,1	-	Ácido vanílico β-glucósido ^①		92776	
27	595,1608	18,24	C 27 H 31 O 15	3,8	9,8	-	Naringenina dihexósido ^①	347846	921102	227723
A	757,1608	18,41	C 35 H 33 O 19	0,3	59,9	-	Desconocido		176435	
28	327,1057	18,48	C 15 H 19 O 8	10,2	17,2	-	3,4'-Dihidroxipropiofenona-3-β-D-glucósido	89840		68284
B	343,1356	18,82	C 16 H 23 O 8	0,5	2,2	-	Desconocido		1552019	
29	385,1837	19,15	C 19 H 29 O 8	-6,3	20,2	-	Vomifoliol β-D-glucopiranosido	104823	282567	52075
30	453,1052	20,04	C 20 H 21 O 12	-1,6	3,9	-	1-O-β-D-(6'-O-Galoil)glucopiranosil-3-metoxi-5-hidroxibenceno ^①		1691557	
C	341,1220	19,29	C 16 H 21 O 8	6,9	2,9	-	Desconocido		1881012	
31	457,0738	20,50	C 22 H 17 O 11	14,9	19,8	-	(-)-Epigalocatequina galato	311407	979819	396667
32	327,1417	20,92	C 16 H 23 O 7	3,5	6,0	-	Betulósido	8408892	15977041	8068000
33	457,1113	21,80	C 23 H 21 O 10	12,6	23,3	-	Daidzina 6"-O-acetato ^①	1853812	4369759	1717758
D	611,2547	21,92	C 26 H 43 O 16	2,1	10,1	-	Desconocido		359065	
34	479,0779	22,25	C 21 H 19 O 13	0,8	3,6	-	Miricetina hexósido I (isómero)	3357129	7235765	3960995
35	479,0788	22,49	C 21 H 19 O 13	4,5	2,3	-	Miricetina hexósido II (isómero)	1539125	3005835	1613762
E	495,1487	22,62	C 23 H 27 O 12	5,0	11,0	-	Desconocido		378083	
36	595,1238	22,86	C 26 H 27 O 16	5,7	12,6	-	Quercetina diglicósido	273922	701364	193556
37	449,0683	23,11	C 20 H 17 O 12	2,9	13,7	-	Miricetina 3-arabinósido I (isómero)	158751	371934	155763
38	433,0397	23,63	C 19 H 13 O 12	1,3	4,6	-	Ácido elágico-7-xilósido	978612	1852009	880306
39	449,0719	23,94	C 20 H 17 O 12	1,3	2,2	-	Miricetina 3-arabinósido II (isómero)	1358646	3595339	1331687
40	449,0689	24,37	C 20 H 17 O 12	3,2	6,7	-	Miricetina 3-arabinósido III (isómero)	1607473	3845340	1715921
41	463,0837	24,68	C 21 H 19 O 12	2,3	4,5	-	Quercetina glucósido I (isómero)	2785871	6184886	2128048
42	463,0838	24,99	C 21 H 19 O 12	2,8	8,3	-	Quercetina glucósido II (isómero)	1047897	2854565	862214
43	523,2113	25,63	C 26 H 35 O 11	2,7	9,1	-	Secoisolariciresinol β-D-glucósido ^①	144060	377298	268276
44	433,0727	25,87	C 20 H 17 O 11	0,7	17,0	-	Quercetina glicósido I (isómero) ^①	367069	1056520	278714
45	433,0733	26,33	C 20 H 17 O 11	2,6	11,7	-	Quercetina glicósido II (isómero) ^①	1379491	3574817	1356858



F	479,1536	26,54	C 23 H 27 O 11	4,7	13,9	-	Desconocido	33144	245841	81027
46	433,0715	26,93	C 20 H 17 O 11	2,6	8,8	-	Quercetina glicósido III (isómero) ^①	1941852	4954426	1705299
47	565,2208	27,31	C 28 H 37 O 12	5,3	7,7	-	Lioniresinol 3 α -O-L-ramnósido ^①	137675		145246
G	537,1603	28,26	C 25 H 29 O 3	0,8	9,9	-	Desconocido	169923	369769	138390
48	289,1069	28,76	C 16 H 17 O 5	12,9	48,5	-	Isosibiricina ^①	143241	64258	
49	435,1173	29,33	C 21 H 23 O 10	6,4	25,6	-	Catequina 3-O- α -L-ramnopiranosido	214327	457027	118748
50	317,0272	29,42	C 15 H 9 O 8	10,7	23,9	-	Miricetina		520694	
51	405,1189	29,85	C 20 H 21 O 9	2,7	3,6	-	Tetrahydroestilbeno glucósido ^①	884437	1583853	730208
H	419,0974	30,93	C 20 H 19 O 10	3,1	3,6	-	Desconocido	465484	1143508	604583
52	609,1234	31,59	C 30 H 25 O 14	1,2	15,3	445,2071	Quercetina cumaroil-hexósido I (isómero) ^①	77652	381395	209865
53	609,1245	31,97	C 30 H 25 O 14	1,3	7,1	-	Quercetina cumaroil-hexósido II (isómero) ^①	325922	1024248	1942534
54	609,1247	32,28	C 30 H 25 O 14	5,2	6,2	-	Quercetina cumaroil-hexósido III (isómero) ^①		99508	90628
55	593,1195	33,95	C 30 H 25 O 13	2,7	4,4	-	Kaempferol diglucósido I (isómero)	796265	2279245	868884
56	609,1218	33,18	C 30 H 25 O 14	4,7	12,0	-	Quercetina cumaroil-hexósido IV (isómero) ^①		150884	240455
I	429,1037	32,75	C 18 H 21 O 12	0,4	21,4	-	Desconocido		131192	
57	301,0331	34,41	C 15 H 9 O 7	7,6	6,3	-	Quercetina ^①		392322	
58	593,1192	34,69	C 30 H 25 O 13	4,5	5,7	-	Kaempferol diglucósido II (isómero)	249665	829600	286276
J	327,2159	36,63	C 18 H 31 O 5	4,5	13,5	-	Desconocido		59742	

S2, extracto acuoso, spray drying.

S4, extracto acuoso de material recolectado en agosto-septiembre, spray drying.

S2a, fracción polifenólica derivada de S2.

^① Compuestos descritos en *Cistus* por primera vez



Para un total de 68 picos, se determinaron las formulas moleculares y se identificaron 58 compuestos, 18 de los cuales son descritos por primera vez en *Cistus*. En la Tabla 30 se detalla la masa exacta, el tiempo de retención, la fórmula molecular con los valores de error y $mSigma$ correspondientes y los fragmentos detectados en el espectro de masas de algunos compuestos. El patrón de fragmentación de los compuestos 20, 21 y 23 es atípico, así como la abundancia de los iones principales. Esto se debe a la disociación en la fuente inducida por colisión (ISCID). La intensidad mostrada en la tabla para estos picos fragmentados en la fuente de ionización, corresponde con la intensidad de los fragmentos mayoritarios. La masa exacta y otros datos específicos para el pico 21 no están disponibles debido a que el rango de masas del método fue menor que la masa de este compuesto (de 50 a 1100 m/z). De cualquier manera, estudios previos han confirmado su correcta identificación⁵¹.

Los compuestos más abundantes en todas las muestras de *C. salviifolius* fueron los ácidos carboxílicos (ácidos quínico y shikímico), los flavanoles (catequinas), el betulósido, los elagitaninos (punicalaginas) y flavonoles glicosilados (glicósidos de miricetina y quercetina). Sin embargo, se observaron diferencias significativas entre los perfiles de S2, S2a y S4. Aunque se encontraron 34 compuestos en común entre los tres extractos, hasta 52 compuestos coincidieron en los extractos de S2 y S4, lo cual era de esperar dado que ambos derivan de la misma plantación.

Es importante destacar que la purificación del extracto S2 resultó en un decrecimiento significativo del número de compuestos polares tales como ácidos carboxílicos, ácidos fenólicos, galotaninos y derivados del hexahidroxidifenil. De esta forma, no se detectó ningún compuesto entre 0 y 15 minutos del cromatograma del extracto S2a. Aun así se identificaron 50 compuestos en este extracto. Por otra parte, se incrementó significativamente la



cantidad de compuestos hidrofóbicos en el extracto de S2a, como flavonoles glicosilados y algunos flavanoles, comparado con los extractos de S2 y S4.

2.2.3. Aplicación de la herramienta *clustering* de líneas a la caracterización pormenorizada de los extractos de *Cistus* S2, S2a y S4.

Para realizar una comparación entre la caracterización por el método convencional y la realizada con los datos tratados mediante el *clustering* de líneas, se retomó el estudio de los extractos más bioactivos de la colección de *Cistus*. Mediante la utilización de una plataforma analítica distinta a la del estudio anterior, UPLC-ESI-QTOF-MS, se analizaron nuevamente los extractos S2, S2a y S4. Para cada uno de los extractos, se realizaron análisis de MS y análisis de MS/MS automáticos con la finalidad de obtener mayor información estructural. En la Figura 94, se presenta el espectrograma de baja resolución (1s, 1Da) del extracto S4 de *C. salvifolius*. En la Figura 95, se presentan dos zonas a las que se ha aplicado un zoom, correspondientes al rango de masas en el que se encuentran los elagitaninos. En la gráfica de la parte superior de esta última figura destacan los dos isómeros del compuesto punicalagina de m/z 1083, detectados a 13 y 15,5 min, respectivamente. La gráfica de la parte inferior corresponde con la zona ampliada del espectrograma de arriba de la misma figura. En ésta, se aprecia mejor el segundo isómero de la punicalagina.



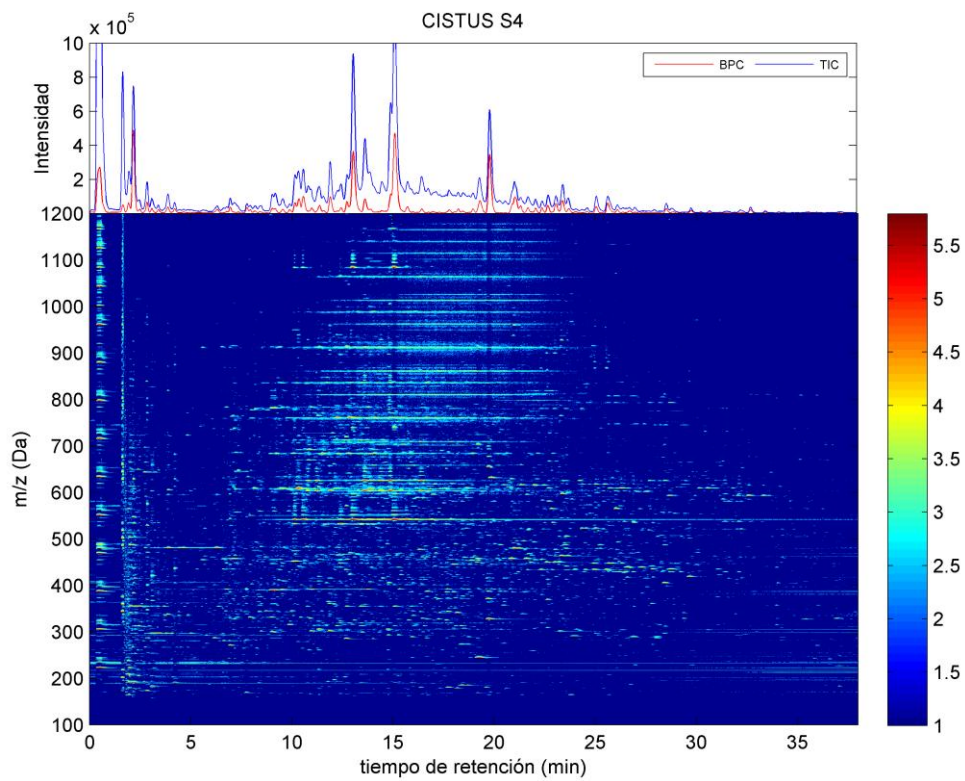


Figura 94. Espectrograma de baja resolución (1 s, 1 Da) del extracto acuoso de *C. salviifolius* (S4) analizado mediante UPLC-QTOF-MS. Con el mapa de color se representa la intensidad para cada par $(t_R, m/z)$.

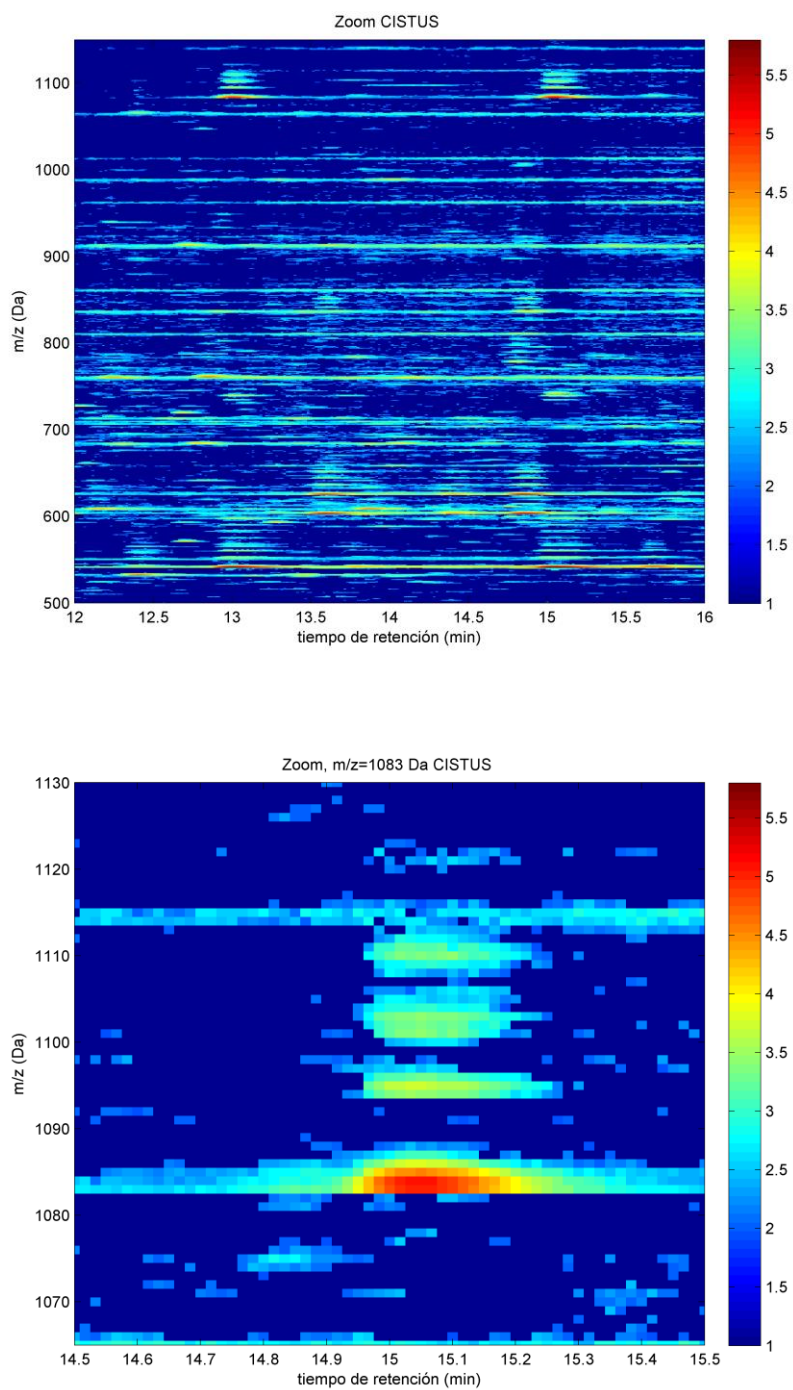


Figura 95. *Zoom* del espectrograma de baja resolución (1 s, 1 Da) realizado con el extracto S4. Con el mapa de color se representa la intensidad para cada par (t,m) .



Para aplicar el método de *clustering*, primero se realizó el preprocesamiento basado en la conversión de los análisis a formato binario, la eliminación de ruido y la calibración en masa. Posteriormente, se aplicó el método *clustering* de líneas para obtener la lista de compuestos. En la Tabla 31 se presentan los datos referentes a los ficheros de las tres muestras analizadas y en la Tabla 32 se muestran los parámetros utilizados en el preprocesamiento.

Tabla 31. Datos de los ficheros de las muestras de *Cistus* analizadas para la caracterización.

Variedad	Ficheros .baf	Ficheros .bin	puntos iniciales($t_R, m/z, I$) del fichero .bin
S2	490.7 MB	2.8 MB	225.260
S2a	492.0 MB	4.8 MB	474.861
S4	492.5 MB	3.1 MB	271.795

Tabla 32. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de *Cistus*.

Ficheros	S2, S2a, S4
Rango de tiempo	[0 s, 38 min] ó [0 s, 2280 s]
Rango de masa	[0 Da, 1200 Da]
Eliminación ruido iónico y Umbral de intensidad	umbral_ruido=80 cuentas
Calibrante	formiato de Na

Tras la aplicación del preprocesamiento de los análisis de *Cistus* seleccionados (S2, S2a y S4), éstos se trataron mediante el *clustering* de líneas, utilizando los parámetros de la Tabla 33.



Tabla 33. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de *Cistus*.

Nombre del parámetro	Valor utilizado
Umbral intensidad	100 cuentas
Umbral área	500
Resolución en tiempo fase 1	300 s
Resolución en tiempo fase 2	20 s
Umbral tiempo de retención	20 s
Umbral m/z	0.010 Da

En la Figura 96 se representan los *clusters* obtenidos, tomando como ejemplo uno de los extractos, el S4. La Figura 97 muestra dos zonas ampliadas con el *zoom*. Se ha representado cada *cluster* como una elipse, en la que un semieje corresponde con la desviación estándar de la masa y el otro, con la desviación estándar del tiempo de retención. En la imagen inferior, se amplía la zona correspondiente al primer isómero de punicalagina de masa m/z 1083, pudiendo apreciarse la forma elíptica en la que se representa el *cluster* del mismo mediante la herramienta utilizada.

En el apéndice 2, se presentan las tablas con los datos correspondientes a los *clusters* generados en cada fase del *clustering*, el tiempo computacional y el histograma del área de los *clusters* de la muestra S4. También se ha incluido el listado de compuestos identificados mediante el *clustering* de líneas para un umbral determinado de área. En esta lista, se han marcado en color gris los compuestos identificados que coinciden con los descritos en la caracterización realizada con el método convencional.



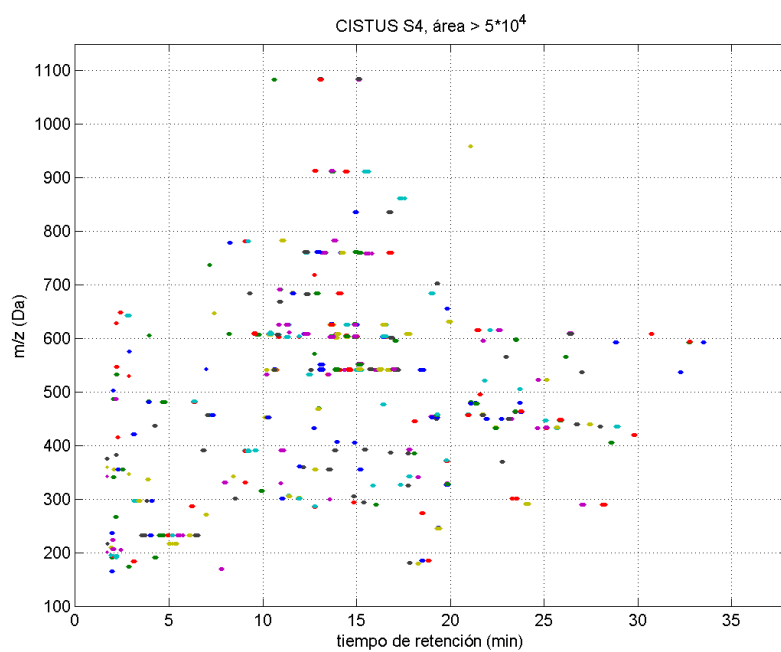


Figura 96. *Clusters* generados con el método de *clustering* de líneas a la muestra de *Cistus* S4.

Se ha aplicado un umbral de área de 10^4 cuentas.



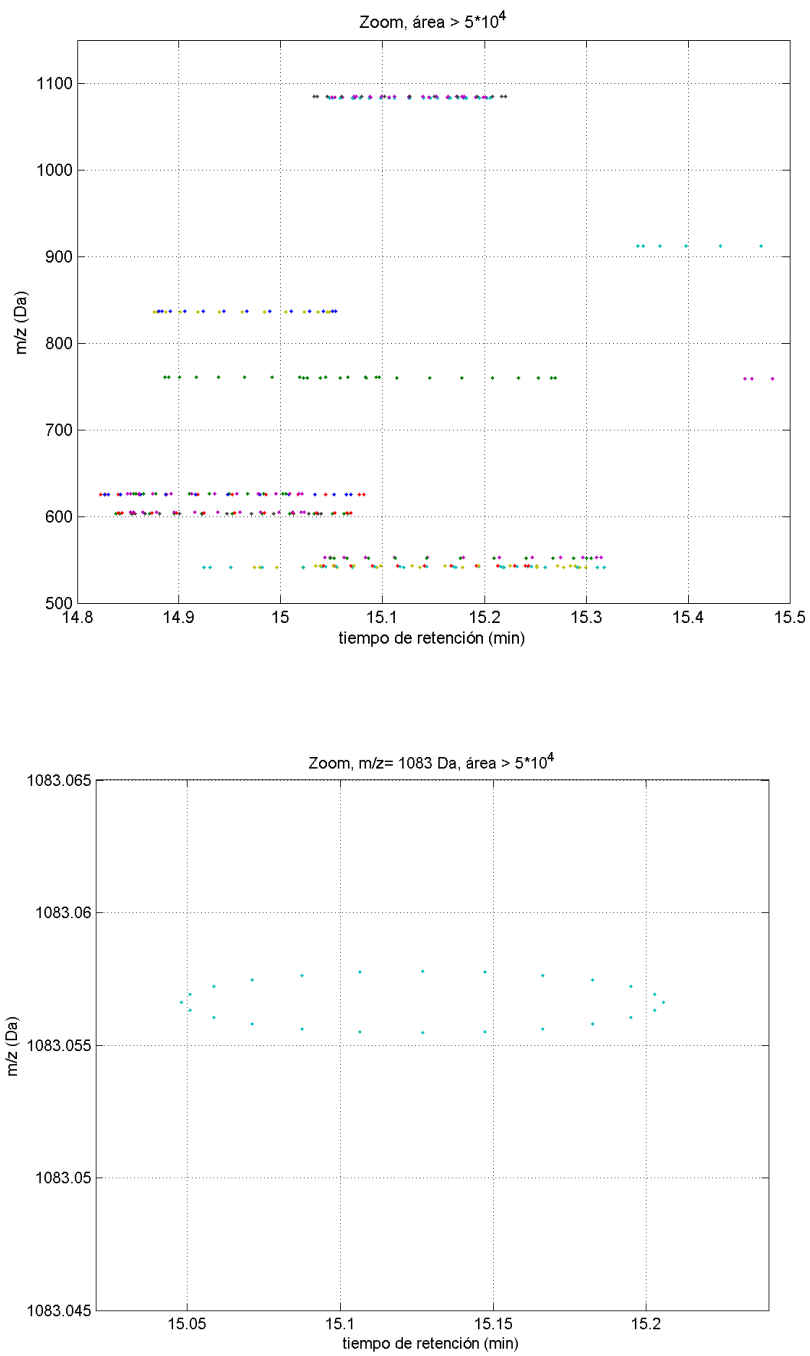


Figura 97. Zoom de la representación de los *clusters* obtenidos para el extracto S4 tras aplicar la herramienta de *clustering* de picos con un umbral de área de 5×10^4 .



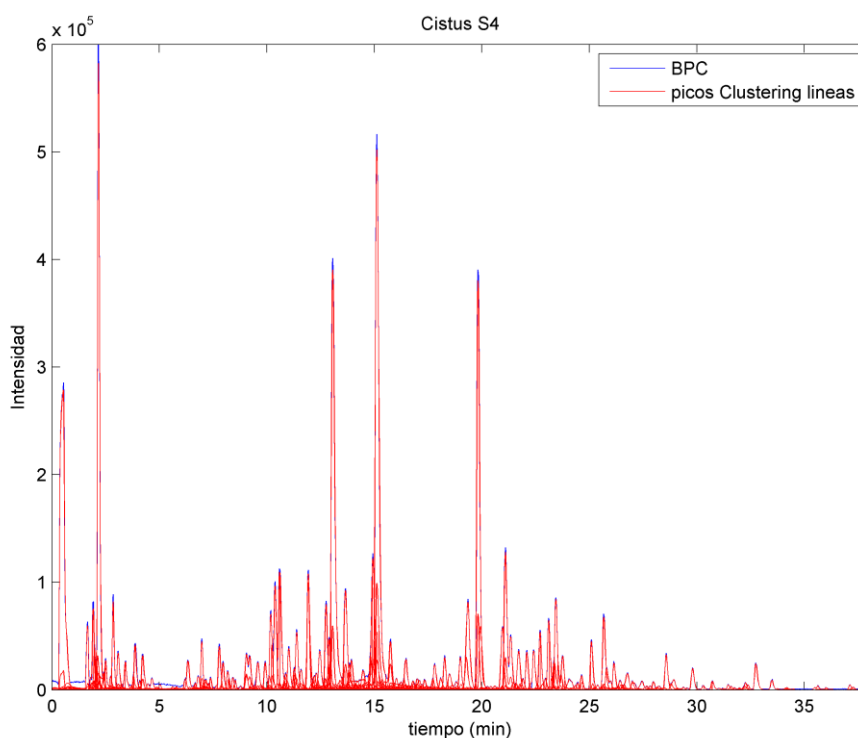


Figura 98. Cromatograma BPC del *Cistus S4* (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (en rojo).

En la Figura 98 se ha representado el cromatograma BPC (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área mayor de 1×10^4 cuentas (en rojo) para la muestra de *Cistus S4*. Mediante el *clustering* se han identificado 536 picos correspondientes a los representantes de las familias de líneas. Los picos minoritarios que coeluyen con los mayoritarios y que permanecen ocultos se pueden apreciar mejor en la figura 2, donde se muestran detalles de diferentes zonas del cromatograma.

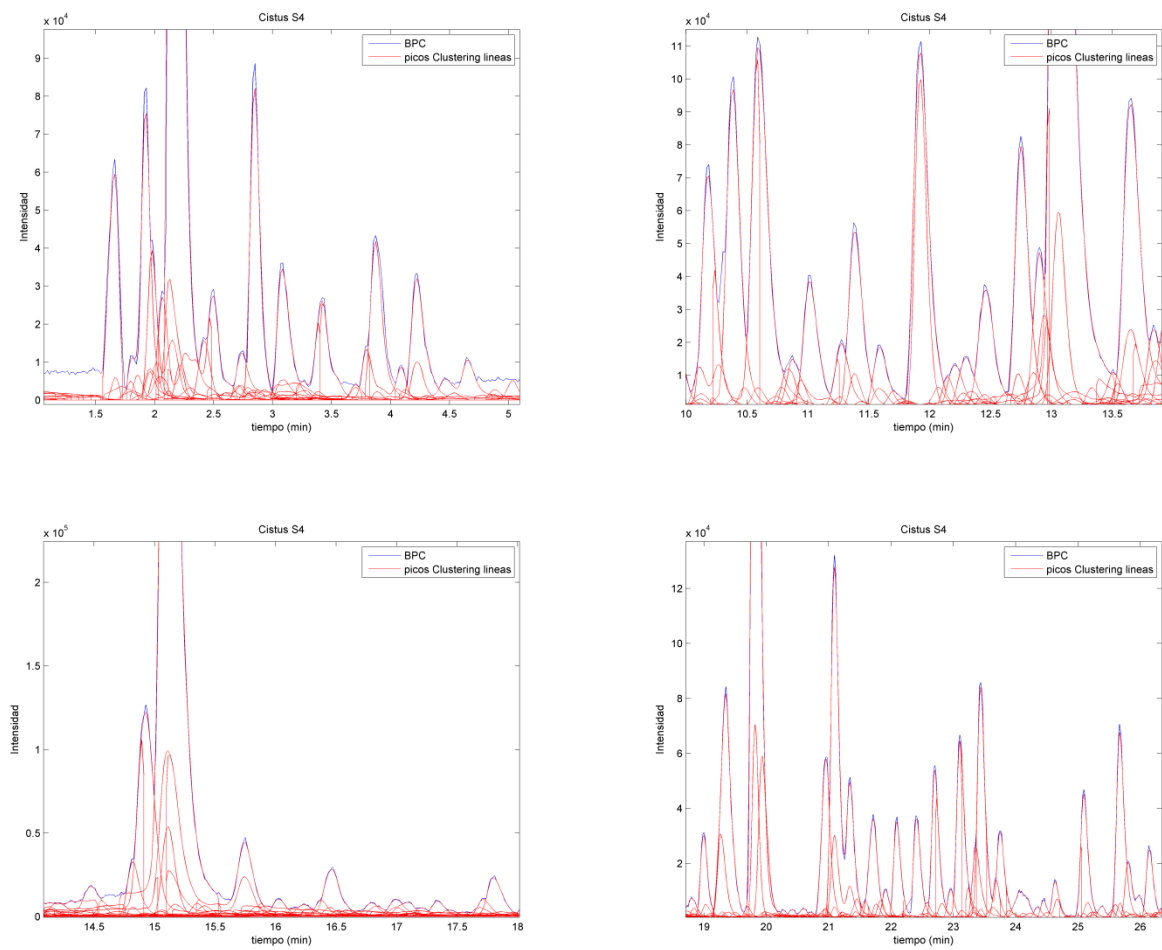


Figura 99. Diferentes detalles de la Figura 98. BPC (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (en rojo).



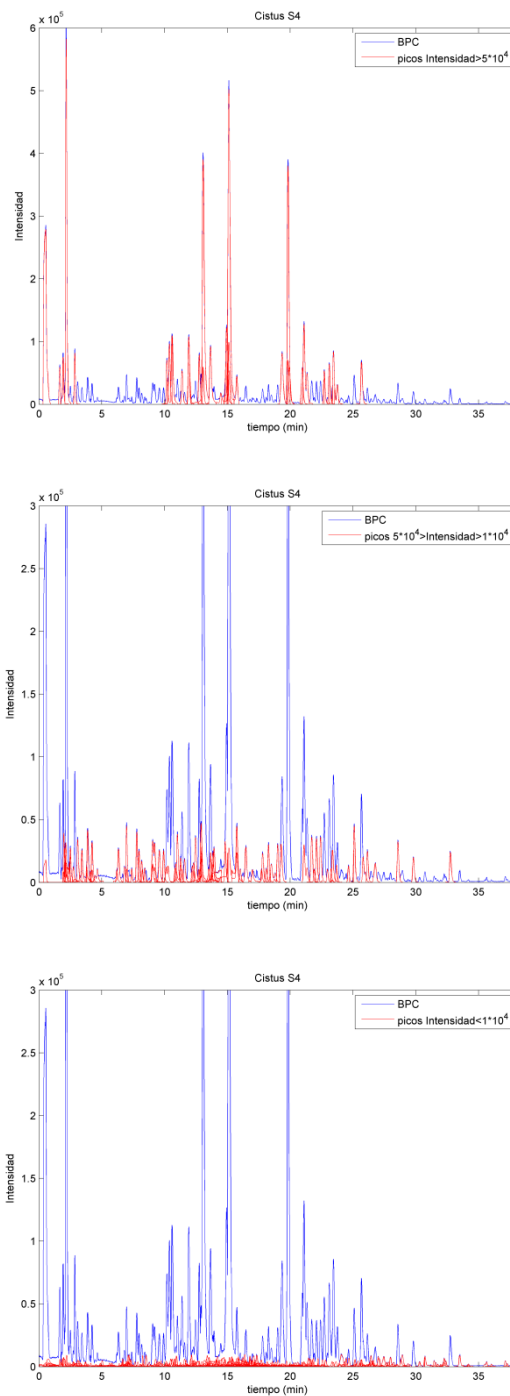


Figura 100. Cromatograma BPC del *Cistus S4* (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (picos en rojo). Se han representado tres gráficas correspondientes a los picos con intensidad $> 5 \times 10^4$ cuentas (35 picos), los picos con intensidad comprendida entre 5×10^4 y 1×10^4 cuentas (95 picos) y los picos con intensidad $< 1 \times 10^4$ cuentas (406 picos).



En la Figura 100 se muestran tres gráficas correspondientes a los picos con intensidad mayor de 5×10^4 cuentas (35 picos), los picos con intensidad comprendida entre 5×10^4 y 1×10^4 cuentas (95 picos) y los picos con intensidad de menor de 1×10^4 cuentas (406 picos). Como se aprecia, el número de picos detectado mediante el *clustering* es mucho más elevado que el número de picos que se pueden identificar mediante el método convencional ya que el número de picos que coeluyen y aparecen ocultos es muy elevado. El siguiente paso consiste en identificar cada uno de los picos encontrados mediante el *clustering*.

En este estudio de caracterización de extractos de pimiento, se ha analizado la lista generada por el programa *clustering* de líneas, comprobando que se obtienen todos los compuestos identificados en el trabajo de caracterización convencional. Además, la aplicación de la herramienta ha permitido la identificación de 17 compuestos más, que se suman a los 58 compuestos identificados en el estudio anterior. De estos compuestos nuevos, se ha de destacar que 10 se encontraban ocultos debajo de otros picos más intensos y que 12 son descritos en *Cistus* por primera vez. Para permitir una mejor apreciación de los compuestos nuevamente identificados, se presentan en la Figura 101 sus cromatogramas de ion extraído bajo el cromatograma de pico base, aplicando el zoom necesario en cada caso. Los números de pico indicados en los cromatogramas coinciden con los de la Tabla 34, donde se detallan los resultados de la caracterización pormenorizada de los extractos de *Cistus* S2, S2a y S4 aplicando el *clustering* de líneas.

Para cada compuesto identificado, la Tabla 34 incluye la información de fragmentación e intensidad relativa de los fragmentos, obtenida en los análisis de MS/MS automáticos. Dichos análisis proporcionan información estructural complementaria que ha resultado de gran utilidad para la caracterización de los compuestos de la lista proporcionada por el *clustering* de líneas.



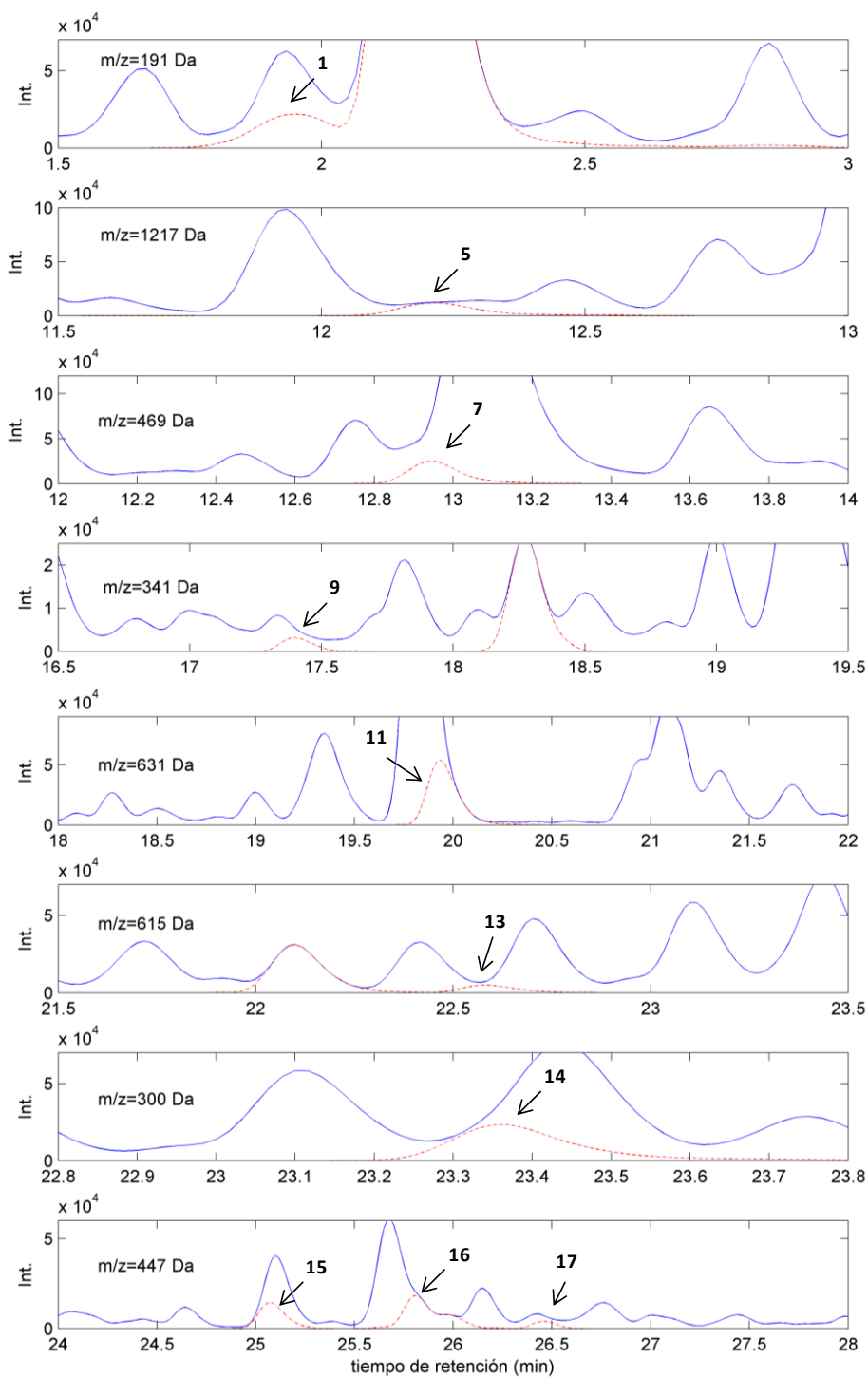


Figura 101. Cromatogramas de ion extraído (EIC) bajo el cromatograma de pico base (BPC) de los picos ocultos identificados gracias a la aplicación del *clustering* de picos. Corresponde con el extracto S4.



Tabla 34. Caracterización de los extractos de *Cistus salviifolius* S2, S2a y S4 mediante el tratamiento de los datos con el *clustering* de líneas.

#	RT (min)	Experimental [M-H] ⁻	Calculada [M-H] ⁻	Fórmula molecular [M-H] ⁻	Error (ppm)	mSigma	MS ² fragmentos (Intensidad relativa %)	Compuesto	Área		
									S2	S2a	S4
1 *	1.93	191.0186	191.0197	C ₆ H ₇ O ₇	6.1	8.6	No fragmentado	Ácido cítrico isómero I ^①	155467	-	129587
2	4.22	191.0183	191.0197	C ₆ H ₇ O ₇	7.6	11.3	No fragmentado	Ácido cítrico isómero II ^①	487026	-	275739
3	9.92	315.0722	315.0722	C ₁₃ H ₁₅ O ₉	-0.1	5.0	No fragmentado	Gentisoil hexósido	122338	-	204013
4	11.91	301.0928	301.0929	C ₁₃ H ₁₇ O ₈	0.4	5.0	No fragmentado	β-D-Glucopiranosido, hidroxí-(hidroximetil) fenil ^① o β-D-Glucopiranosido, hidroxí-metoxifenil	4939	1343149	993946
5 *	12.23	1217.2400	1217.2416	C ₆₀ H ₄₉ O ₂₈	1.3	27.3	608.1132 (100.0) 912.1502 (10.5)	GC4 (tetrámero) ^①	132381	255709	109130
6	12.45	1065.1938	1065.1942	C ₅₂ H ₄₁ O ₂₅	0.5	16.1	532.0918 (100.0) 760.1306 (9.0) 897.1536 (2.3)	GC2-GCG (trímero) ^①	283064	371736	342238
7 *	12.95	469.0040	469.0049	C ₂₁ H ₉ O ₁₃	1.9	12.3	425,0158 (100.0) 407,0073 (5,1) 300,9963 (0,7)	Ácido flavogalónico ^①	1481393	102337	236831
8	16.79	387.1655	387.1661	C ₁₈ H ₂₇ O ₉	1.3	12.5	No fragmentado	Ácido tuberónico glucósido ^① o Ácido 5'-Hidroxijasmónico 5'-O-glucósido	190843	421680	73267
9 *	17.39	341.1252	341.1242	C ₁₆ H ₂₁ O ₈	-2.9	17.0	179,0700 (100,0) 180,0779 (5,0)	Coniferina isómero I ^①	35235	98820	28406



10	18.26	341.1246	341.1242	C ₁₆ H ₂₁ O ₈	-1.2	6.2	179,0701 (100,0) 180,0735 (11,7)	Coniferina isómero II ^①	413484	1005923	245000
11 *	19.94	631.0915	631.0941	C ₂₈ H ₂₃ O ₁₇	4.1	1.1	479.0842 (14,3)	Miricetina 3-galoilhexósido ^①	477846	1248344	2626
12	22.10	615.0986	615.0992	C ₂₈ H ₂₃ O ₁₆	0.8	4.5	463,0880 (17,8) 301,0295 (1,7)	Quercetina glicósido galato isómero I ^①	275040	686289	297295
13 *	22.58	615.0985	615.0992	C ₂₈ H ₂₃ O ₁₆	1.1	14.8	463.0841 (14,8)	Quercetina glicósido galato isómero II ^①	47874	132667	43943
14 *	23.35	300.9993	300.9990	C ₁₄ H ₅ O ₈	-1.2	3.9	No fragmentado	Ácido elágico	173352	378260	249127
15 *	25.06	447.0927	447.0933	C ₂₁ H ₁₉ O ₁₁	1.3	7.2	284,0325 (70,1) 285,0405 (30,7)	Kaempferol glicósido isómero I	136118	282294	144520
16 *	25.88	447.0926	447.0933	C ₂₁ H ₁₉ O ₁₁	1.5	7.9	284,0338 (54,7) 285,0405 (41,1)	Kaempferol glicósido isómero II	132957	282973	168092
17 *	26.47	447.0563	447.0569	C ₂₀ H ₁₅ O ₁₂	1.3	25.8	315,0151 (100,0)	Ácido elágico-7-xilósido metilado	27519	65477	37676

* Compuestos ocultos bajo otros picos de mayor intensidad.

^① Compuestos descritos por primera vez en *Cistus*.

(-): compuesto no detectado.

GC: (epi)galocatequina.

GCG: (epi)galocatequina galato.



Algunos de los compuestos, principalmente, aquellos de menor masa molecular, no han sido fragmentados en el análisis de MS/MS automático, por lo que su identificación se basa en el espectro de masas, así como en su identificación previa según otros autores.

Los picos 1 y 2 se han identificado como isómeros del ácido cítrico, eluyendo a 1.93 y 4.22 min, respectivamente. Ambos son descritos en *Cistus* por primera vez, encontrándose el primero de ellos oculto bajo otro pico de mayor intensidad²¹⁹. El pico 3 de m/z 315.0722 se ha identificado como gentisoil hexósido, previamente descrito en *Cistus ladanifer*²²⁰. Los picos 1, 2 y 3 no se han identificado en el extracto S2a dado que en el cromatograma de este extracto purificado no se detectan compuesto hasta aproximadamente el minuto 10.

No se han encontrado fragmentos para el pico 4 (m/z 301.0928), por lo que la identificación de su fórmula molecular C₁₃H₁₈O₈ se propone por primera vez en *Cistus* como hidroxi-(hidroximetil)fenil-β-D-glucopiranosido o hidroximetoxifenil-β-D-glucopiranosido²²¹.

Los picos 5 y 6 se identificaron como tetrámero de galocatequina y trímero de galocatequina, respectivamente. Ambos son picos ocultos bajo otros de mayor intensidad y descritos en *Cistus* por primera vez. En la Figura 102 se muestra la estructura del compuesto 5, indicando los fragmentos observados en el espectro de MS/MS. Para este compuesto de m/z 1217,2400, el fragmento de mayor intensidad que se observa es a m/z 608.1132, correspondiente con la

²¹⁹ Liu, R.H., 2003. Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. *The American Journal of Clinical Nutrition*, 78, p.517S–520S.

²²⁰ Barros, L. *et al.*, 2013. Antifungal activity and detailed chemical characterization of *Cistus ladanifer* phenolic extracts. *Industrial Crops and Products*, 41, pp.41–45.

²²¹ Maldini, M. *et al.*, 2009. Phenolic compounds from *Bursera simaruba* Sarg. bark: phytochemical investigation and quantitative analysis by tandem mass spectrometry. *Phytochemistry*, 70(5), pp.641–649.



mitad de la molécula (dímero de epigalocatequina). El fragmento a m/z 912.1502 correspondería con la pérdida de una unidad de epigalocatequina²²².

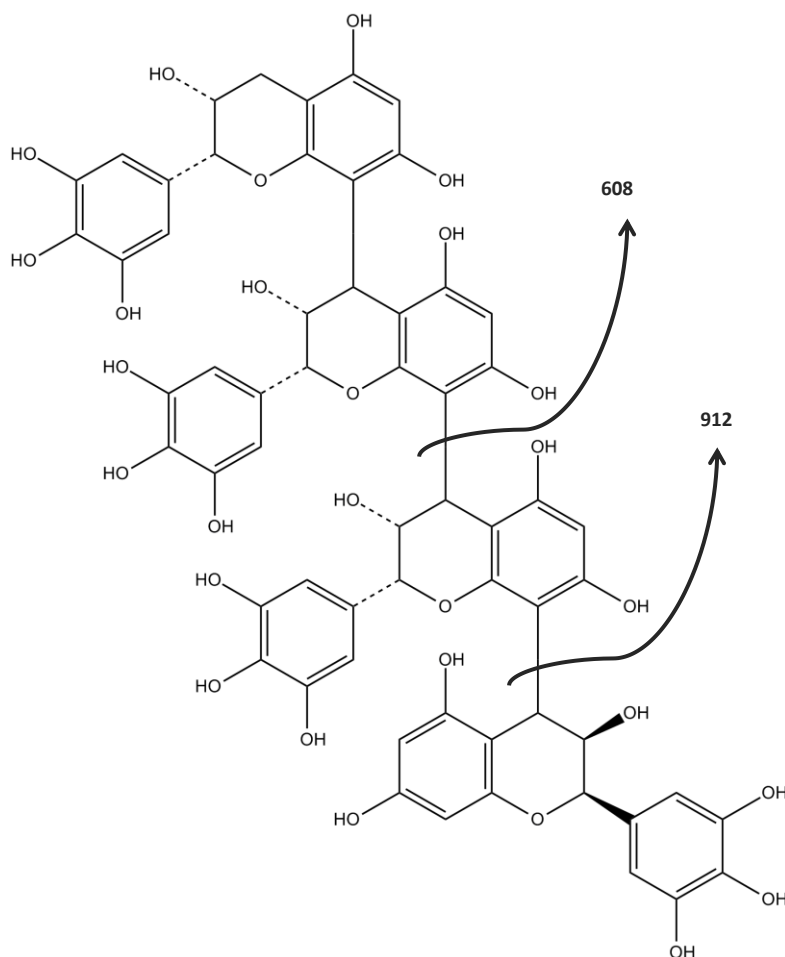


Figura 102. Patrón de fragmentación para el compuesto tetrámero de galocatequina (GC4).

El pico 6 es un trímero compuesto por dos epigalocatequinas y una epigalocatequina galato. El fragmento principal observado de m/z 532.0918 corresponde con la mitad de la molécula. El fragmento a 760.1306 ([M-305-H]) podría deberse a la escisión del enlace interflavonoide, mientras que el fragmento minoritario a m/z 897.1536 ([M-168-H]) podría deberse a la ruptura del sistema cíclico por la reacción inversa a Diels-Alder, conocida como retro-Diels-Alder²²².

²²² Wang, H. *et al.*, 2012. Profiles and α -Amylase Inhibition Activity of Proanthocyanidins in Unripe *Manilkara zapota* (Chiku). *Journal of Agricultural and Food Chemistry*, 60, pp.3098–3104.



El compuesto que eluye a 12.95 min de m/z 469,0040 (pico 7) se ha identificado como ácido flavogalónico. Constituye un compuesto descrito por primera vez en *Cistus*, cuya identificación ha sido posible gracias a que el patrón de fragmentación coincide con el descrito por otros autores en plantas medicinales del género *Terminalia*, entre otros ²²³.

El siguiente compuesto siguiendo el orden de la tabla (pico 8) también es un compuesto nuevo en *Cistus*. Éste no ha sido fragmentado, por lo que su identificación se basa en la fórmula molecular C₁₈H₂₈O₉, proporcionada por su masa exacta (m/z 387,1655) y distribución isotópica. Por ello, este compuesto se ha propuesto como ácido tuberónico glucósido o ácido 5'-Hidroxijasmónico 5'-O-glucósido ²²⁴.

Los picos 9 y 10 de m/z 341,1252 y 341,146, respectivamente se han identificado como isómeros de la coniferina, un glucósido del alcohol coniferílico. Se trata de dos compuestos descritos por primera vez en *Cistus*, encontrándose el primero de ellos eluyendo a 17.39 min por debajo de otro pico de mayor intensidad. Los dos compuestos poseen el mismo patrón de fragmentación, con alguna diferencia entre la intensidad relativa de los fragmentos. El patrón de fragmentación y el espectro de MS/MS de la coniferina se muestra en la Figura 103A. El patrón de fragmentación presentado en la figura corresponde con el primer isómero de la coniferina, si bien la fragmentación es igual para ambos isómeros, salvo diferencias en cuanto a su intensidad relativa.

²²³ Pfundstein, B. *et al.*, 2010. Polyphenolic compounds in the fruits of Egyptian medicinal plants (*Terminalia bellerica*, *Terminalia chebula* and *Terminalia horrida*): characterization, quantitation and determination of antioxidant capacities. *Phytochemistry*, 71(10), pp.1132–1148.

²²⁴ Quirantes-Piné, R. *et al.*, 2010. Characterization of phenolic and other polar compounds in a lemon verbena extract by capillary electrophoresis-electrospray ionization-mass spectrometry. *Journal of Separation Science*, 33(17-18), pp.2818–2827.



El fragmento mayoritario de m/z 179,0700 corresponde con la pérdida de una unidad de glucosa²²⁵.

El pico 11 es un pico oculto de m/z 631,0915, que se describe por primera vez en *Cistus*. Se ha identificado como miricetina 3-galoilhexósido ya que el fragmento de m/z 479.0842 puede deberse a la pérdida del grupo galoil [$M-152-H$]^{- 226,227}.

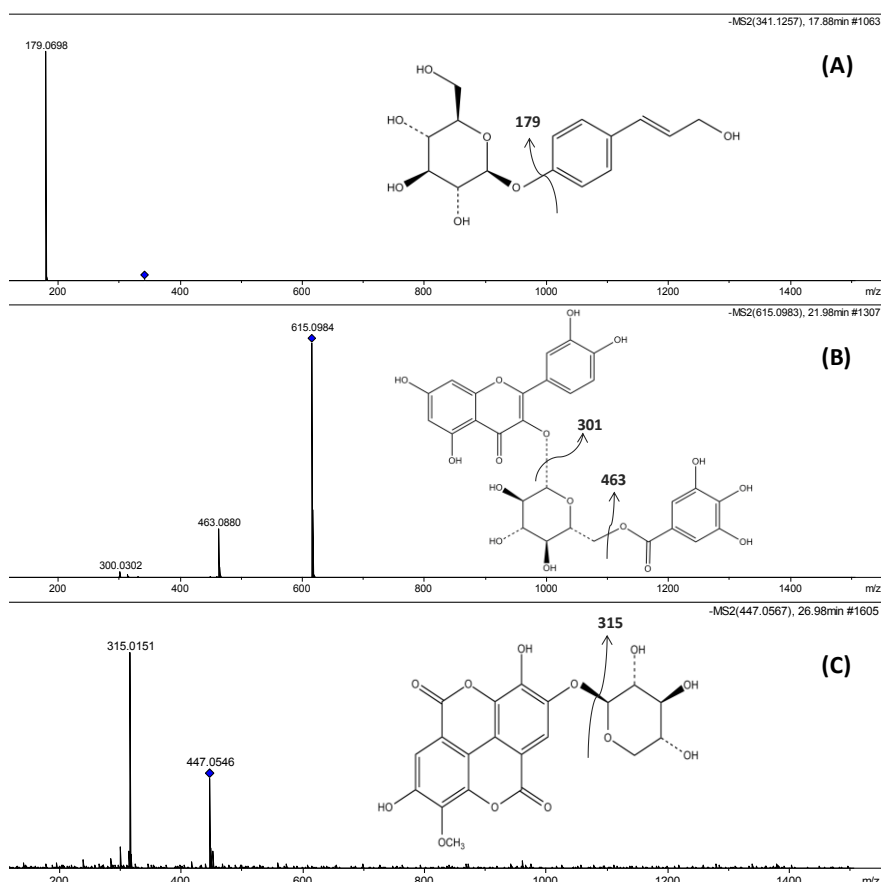


Figura 103. Patrón de fragmentación para (A) coniferina, (B) quercetina glicósido galato y (C) elágico-7-xilósido metilado.

²²⁵ Hu, Y.M. *et al.*, 2010. Quality assessment of *Cortex phellodendri* by high-performance liquid chromatography coupled with electrospray ionization mass spectrometry. *Biomedical chromatography*, 24(4), pp.438–453.

²²⁶ Korul'kina, L.M. *et al.*, 2004. Biologically active compounds from *Limonium*. *Chemistry of Natural Compounds*, 40(5), pp.383–387.

²²⁷ Moharram, F. a *et al.*, 2006. Antioxidant galloylated flavonol glycosides from *Calliandra haematocephala*. *Natural Product Research*, 20(10), pp.927–934.



Los picos 12 y 13 se han identificado como isómeros de la quercetina glicósido galato (con m/z 615,0986 y 615,0985, respectivamente). Se ha de destacar que ambos son descritos por primera vez en *Cistus* y que el segundo isómero se encuentra oculto bajo otro pico de mayor intensidad. En la Figura 103B, se muestra el patrón de fragmentación y el espectro de MS/MS para el pico 12. El fragmento a m/z 301,0295 corresponde con la quercetina, producido por la pérdida del grupo galoil y la deoxihexosa ($[M-152-146-H]^-$). El fragmento de m/z 463,0880 correspondería con la pérdida únicamente del grupo galoil²²⁸.

El compuesto correspondiente al pico 14 ha sido previamente descrito en *Cistus*. Sin embargo, no se había detectado en la caracterización anterior de estos extractos de *Cistus* realizada con el procedimiento convencional debido a encontrarse oculto bajo otro pico de mayor intensidad. Este pico de m/z 300,9993 se ha identificado como ácido elágico¹³⁰.

Los últimos tres picos de la tabla de compuestos identificados son compuestos ocultos bajo picos de mayor intensidad y, por tanto, su identificación se debe a la aplicación del tratamiento de los datos con el *clustering* de líneas. Los picos 15, 16 y 17 poseen una masa de 447, difiriendo la m/z exacta del último.

La identificación de los picos 15 y 16 de m/z 447,0927 y 447,0926, respectivamente, se ha propuesto como isómeros de kaempferol glicósido. El patrón de fragmentación de ambos compuestos desvela la pérdida de la molécula de azúcar¹⁴⁵. Estos compuestos se encuentran descritos en polen de abeja de *Cistus ladanifer*²²⁹.

²²⁸ Rodrigues, C.M. *et al.*, 2008. High-performance liquid chromatographic separation and identification of polyphenolic compounds from the infusion of *Davilla elliptica* St. Hill. *Phytochemical Analysis*, 19(1), pp.17–24.

²²⁹ Tomás-Lorente, F. *et al.*, 1992. Flavonoids from *Cistus ladanifer* from bee pollen. *Phytochemistry*, 31(6), pp.2027–2029.



Por último, el pico 17 de m/z 447,0563 y tiempo de retención 26.47 min, se ha identificado como ácido elágico-7-xilósido metilado (conocido como ducheside A), previamente descrito en *Cistus*. En el espectro de MS/MS de este compuesto se observa un fragmento a m/z 315,0151, que correspondería con la pérdida de la xilosa (Figura 103C)¹³⁰.

2.2.4. Aplicación de la herramienta de correlación a la identificación de compuestos antimicrobianos de extractos polifenólicos de *Cistus*.

Con el fin de identificar el/los compuesto/s responsable/s de la capacidad antimicrobiana de los extractos polifenólicos de *Cistus*, se ha aplicado la herramienta de correlación entre datos de espectrometría de masas y bioactividad, desarrollada en la presente tesis.

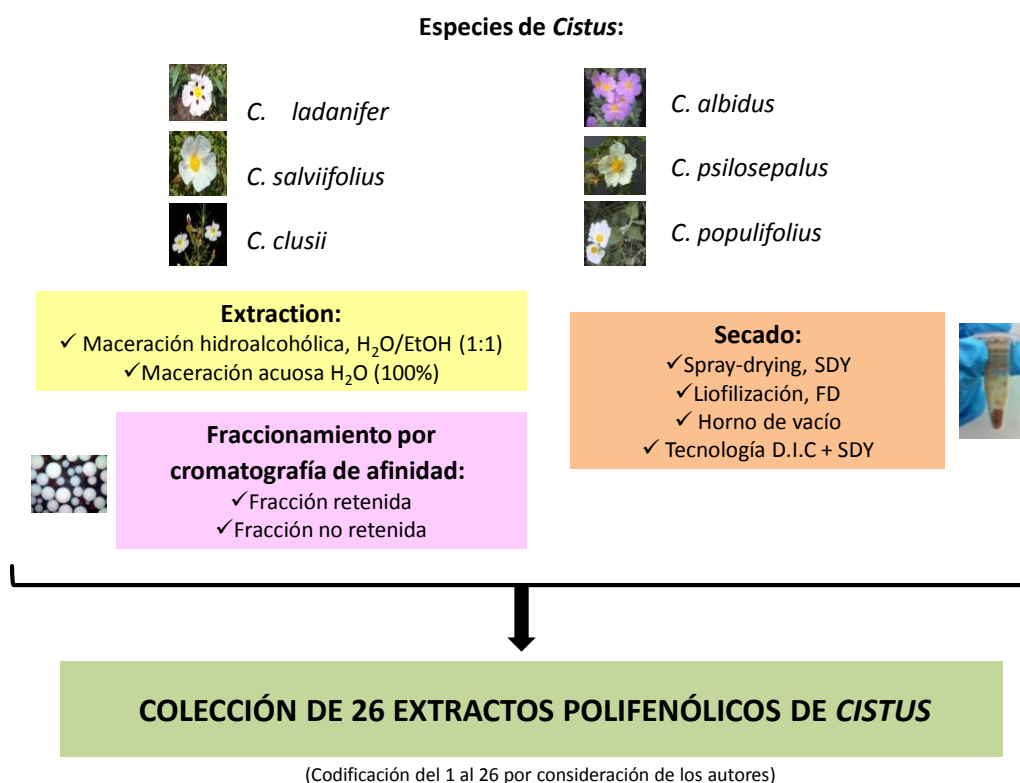


Figura 104. Extractos polifenólicos de *Cistus* utilizados para el estudio de correlación con la bioactividad.

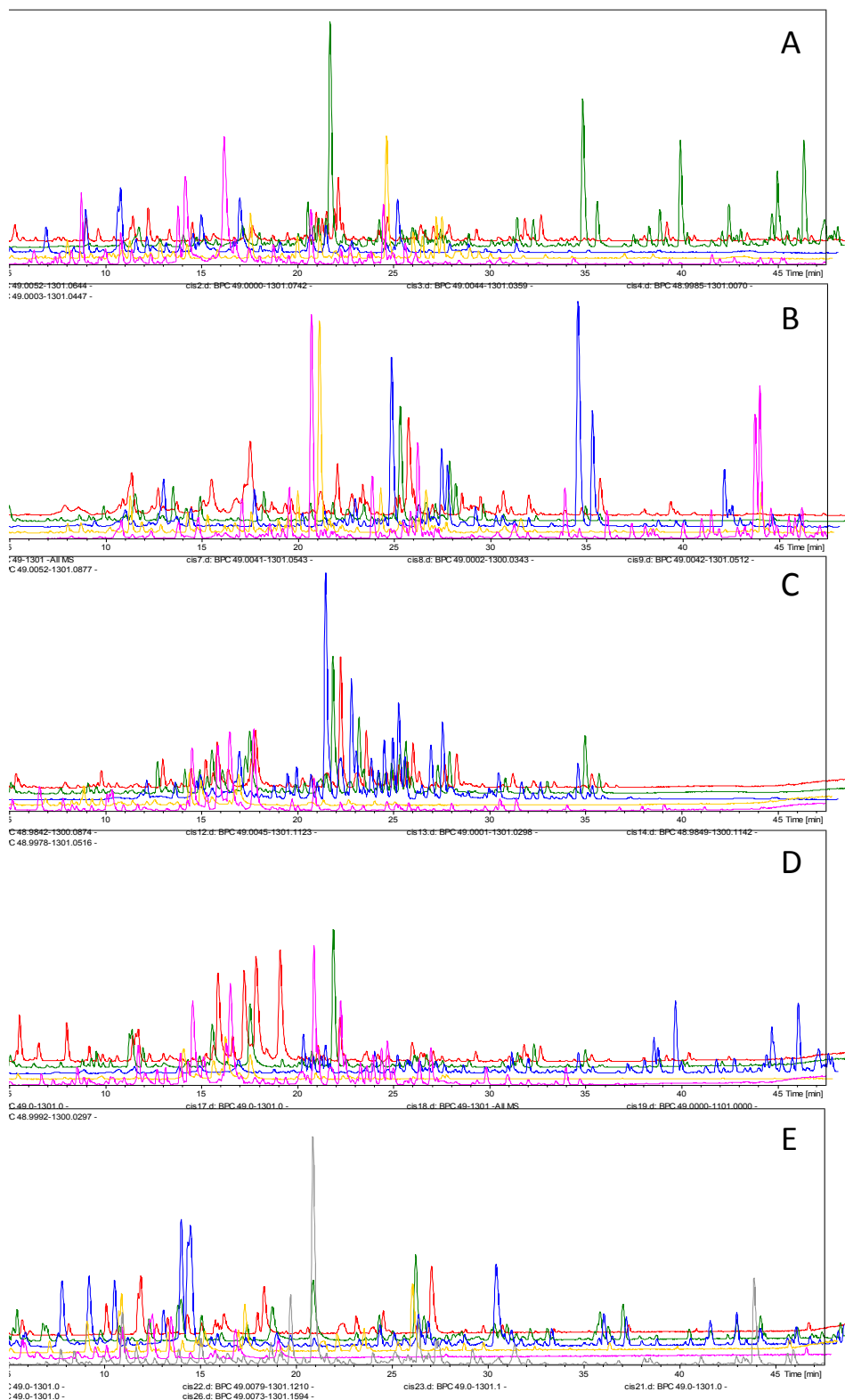


Figura 105. Cromatogramas BPC de los extractos de *Cistus*. A: 1 (rojo), 2 (verde), 3 (azul), 4 (amarillo), 5 (rosa). B: 6 (rojo), 7 (verde), 8 (azul), 9 (amarillo), 10 (rosa). C: 11 (rojo), 12 (verde), 13 (azul), 14 (amarillo), 15 (rosa). D: 16 (rojo), 17 (verde), 18 (azul), 19 (amarillo), 20 (rosa). E: 21 (rojo), 22 (verde), 23 (azul), 24 (amarillo), 25 (rosa) y 26 (gris).



Para ello, se ha ampliado el número de extractos, obteniéndose un total de 26 (Figura 104). La colección de extractos se obtuvo mediante distintas combinaciones entre seis especies diferentes de *Cistus* (*C. ladanifer*, *C. albidus*, *C. clusii*, *C. salviifolius*, *C. psilosepalus* y *C. populifolius*), dos procedimientos de extracción (acuosa e hidroalcohólica), cuatro métodos de secado (liofilización, spray drying, horno de vacío y tecnología D.I.C seguida de spray drying) y un proceso de purificación, utilizando una resina afín a los compuestos fenólicos.

Para obtener los datos analíticos a utilizar en la herramienta de correlación, se analizaron los 26 extractos polifenólicos de *Cistus* mediante HPLC-ESI-TOF-MS. Los perfiles obtenidos se muestran en la Figura 105.

Los datos de bioactividad utilizados para la correlación proceden de ensayos antimicrobianos y se dividen en resultados frente a bacterias Gram positivas y resultados frente a bacterias Gram negativas. En la Figura 106 se representan gráficamente los resultados de actividad antimicrobiana de los 26 extractos de *Cistus*. Para la correlación, se definió un parámetro ($\text{bioactividad} = 1/\text{MIC}_{50}$).

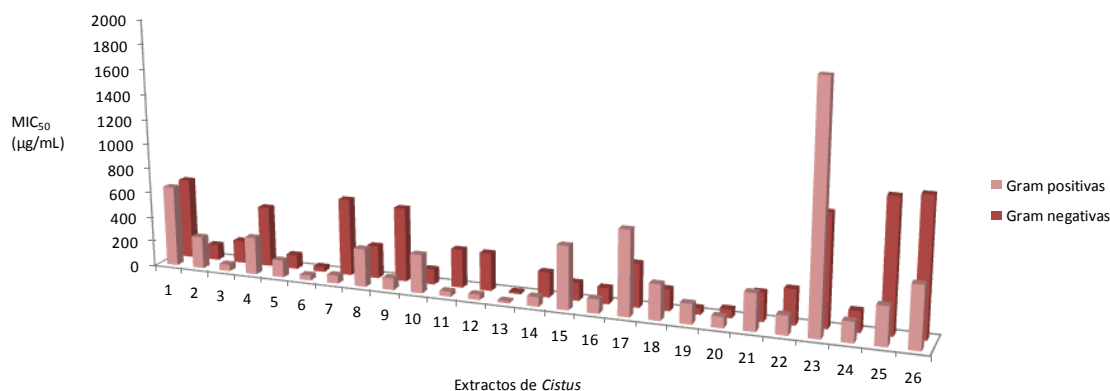


Figura 106. Capacidad antimicrobiana de los 26 extractos de *Cistus* frente a bacterias Gram positivas y Gram negativas.

En la Figura 107 y la Figura 108 se representan los correlogramas realizados con los datos HPLC-MS referentes a los 26 extractos de *Cistus*, previamente tratados con la herramienta *clustering* de líneas y sus correspondientes datos de capacidad antimicrobiana frente a bacterias Gram negativas y Gram positivas, respectivamente. Se representa gráficamente el coeficiente de correlación mediante un mapa de color en función del tiempo-masa de cada *cluster*, de manera que resulta fácil identificar los candidatos puesto que se representan en color rojo, correspondiente a un coeficiente de correlación muy cercano a 1.

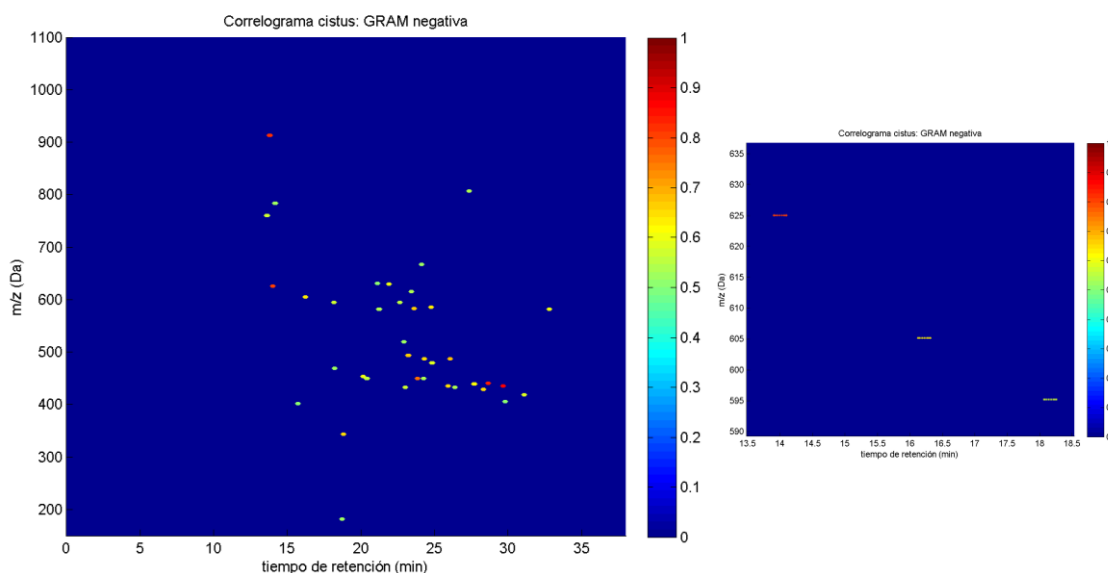


Figura 107. Correlograma realizado con los datos de MS de los 26 extractos de *Cistus* y los datos de capacidad antimicrobiana frente a *E. coli*. En escala de color se representa el coeficiente de correlación para cada *cluster* de los obtenidos con el *clustering* de picos.

Resolución temporal: 1 s y resolución en masa: 1 Da.



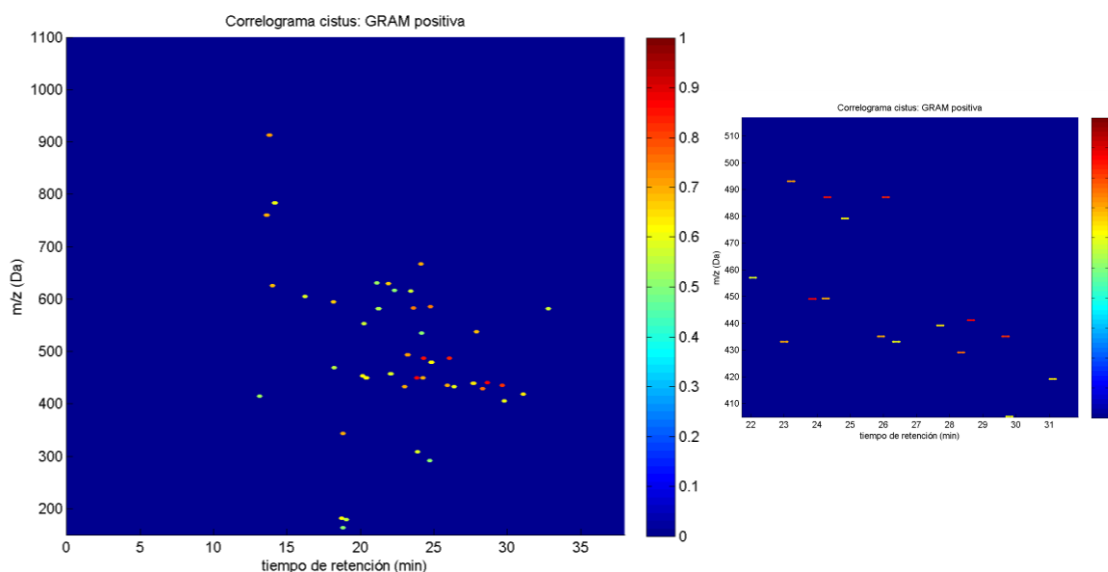


Figura 108. Correlograma realizado con los datos de MS de los 26 extractos de *Cistus* y los datos de capacidad antimicrobiana frente a *S. aureus*. En escala de color se representa el coeficiente de correlación para cada *cluster* de los obtenidos con el *clustering* de picos.

Resolución temporal: 1 s y resolución en masa: 1 Da.

En el zoom de la Figura 107, puede apreciarse uno de los compuestos que ha obtenido un alto coeficiente de correlación en el estudio de bioactividad frente a bacterias Gram negativas. Se trata del *cluster* de m/z 625 Da que aparece a 14 min. En el zoom de la Figura 108, destaca el *cluster* de m/z 449 Da a 23.86 min, el *cluster* de m/z 487 a 24.31 min y un isómero de este último compuesto eluyendo a 26.07 min.



Tabla 35. Listas de compuestos candidatos a ser responsables de la capacidad antimicrobiana de los 26 extractos de *Cistus* frente a bacterias Gram negativas y Gram positivas.

Gram negativa (<i>E. coli</i>)				Gram positiva (<i>S. aureus</i>)			
#	r	m/z (Da)	t _R (min)	#	r	m/z (Da)	t _R (min)
1	0.9325	435.0998	29.68	1	0.9322	449.0724	23.86
2	0.9111	441.1193	28.64	2	0.9279	441.1193	28.64
3	0.9075	913.1783	13.81	3	0.9247	487.1272	24.31
4	0.8953	625.0352	14.00	4	0.9179	487.1289	26.07
5	0.8721	449.0724	23.86	5	0.9094	435.0998	29.68
6	0.8291	487.1289	26.07	6	0.8724	429.1039	28.34
7	0.8131	583.1649	23.63	7	0.8616	583.1649	23.63
8	0.8107	493.1338	23.22	8	0.8573	585.1636	24.77
9	0.8068	343.1386	18.81	9	0.8465	493.1338	23.22
10	0.8052	435.1043	25.93	10	0.8462	913.1783	13.81
11	0.8016	487.1272	24.31	11	0.8378	760.1358	13.62
12	0.8007	429.1039	28.34	12	0.8369	433.0415	23.01
13	0.7995	585.1636	24.77	13	0.8354	343.1386	18.81
14	0.7941	605.1469	16.22	14	0.8328	435.1043	25.93
15	0.7883	439.1057	27.71	15	0.8309	449.2012	24.26
16	0.7788	629.2408	21.92	16	0.8305	667.2575	24.13
17	0.7768	453.1056	20.14	17	0.8288	537.1603	27.92
18	0.7753	581.1846	32.82	18	0.8265	625.0352	14.00
19	0.7675	419.0973	31.10	19	0.8186	629.2408	21.92
20	0.7674	479.1537	24.84	20	0.8112	595.1646	18.16
21	0.7598	433.0415	23.01	21	0.7993	419.0973	31.10
22	0.7594	595.1646	18.16	22	0.7953	439.1057	27.71
23	0.7571	760.1358	13.62	23	0.7881	479.1537	24.84
				24	0.7850	181.0873	18.71
				25	0.7814	453.1056	20.14
				26	0.7814	449.1085	20.41
				27	0.7766	433.0763	26.39
				28	0.7743	783.0671	14.18
				29	0.7713	405.1180	29.80
				30	0.7696	309.1326	23.88
				31	0.7671	457.1169	22.07
				32	0.7637	605.1469	16.22
				33	0.7537	553.1905	20.25
				34	0.7524	581.1846	32.82
				35	0.7504	615.0965	23.44

r: coeficiente de correlación
t_R: tiempo de retención
A: área (en Kcuentas)



Con los parámetros aplicados en la herramienta de correlación ($r > 0.75$), se ha obtenido una lista de 23 candidatos potencialmente responsables de la actividad antimicrobiana Gram negativa y una lista de 35 candidatos potencialmente responsables de la actividad antimicrobiana Gram positiva.

Los coeficientes de correlación obtenidos con los resultados de ambos ensayos antimicrobianos son similares en el sentido de que los coeficientes de correlación más altos son de 0.9325 en Gram negativas y 0.9322 en Gram positivas, correspondientes a los primeros compuestos candidatos de cada lista.

Como se ha comentado anteriormente, los compuestos de alto peso molecular de los extractos de *Cistus* han sufrido fragmentación en la fuente de ionización; por tanto, su huella espectrométrica viene representada por el fragmento de mayor intensidad y su identificación se basa en el patrón de fragmentación en el espectro de masas. Éste es el caso del compuesto de m/z 625.0352 Da a 14 min, que destaca entre los candidatos de la lista de la izquierda (Gram negativas, $r=0.8953$). Se trata de un fragmento de la punicalagina galato, cuya m/z es 1251. Éste mismo compuesto de m/z 625 también aparece como candidato frente a bacterias Gram positivas, pero con un coeficiente de correlación ligeramente menor ($r=0.8265$). Algo similar ocurre con el candidato de m/z 760.1358 Da a 13.62 min que corresponde con un fragmento del trímero de m/z 1065, formado por dos unidades de galocatequina y una unidad de galocatequina galato (GC2-GCG). Este compuesto aparece como candidato en ambas listas, si bien destaca el coeficiente de correlación obtenido con los resultados de bacterias Gram positivas ($r=0.8378$).

El primer candidato de la lista de la derecha (m/z 449.0724 Da a 23.86 min) es la miricetina-3-arabinósido. Este compuesto de la familia de los flavonoles aparece como candidato a compuesto bioactivo frente a ambas cepas, con mayor coeficiente de correlación contra bacterias Gram positivas. El coeficiente de



correlación para este candidato es de $r=0.9322$ contra Gram positivas y $r=0.8721$ contra Gram negativas.

Otros candidatos destacables para el caso de bacterias Gram positivas, son dos isómeros de la vitexina o apigenina glucósido de m/z 487.1272 y 487.1289 y tiempos de retención de 24.31 y 26.07 min, respectivamente²³⁰. Poseen coeficientes de correlación muy altos y similares entre sí ($r=0.9247$, el primer isómero y $r=0.9179$, el segundo). Únicamente uno de los isómeros de este compuesto de la familia de las flavonas, el que eluye a 26.03 min, aparece también en la lista de candidatos a ser responsables de la actividad antimicrobiana Gram negativa, aunque con un coeficiente de correlación menor ($r=0,8291$).

El compuesto de m/z 433.0415 a 23.01 corresponde con el ácido elágico-7-xilósido, siendo candidato bioactivo contra Gram positivas ($r=0.8369$) y Gram negativas ($r=0.7598$).

El candidato de m/z 435.1043 Da que eluye a 25.93 es la catequina 3-O- α -L-ramnopiranosido, con un coeficiente de correlación de 0.8328 para Gram positivas y 0.8052 para Gram negativas.

Entre los candidatos de baja correlación, se ha identificado el 1-O- β -D-(6'-O-Galoil)glucuropiranosil-3-metoxi-5-hidroxibenceno, tanto en la lista de Gram positivas ($r=0.7814$) como en la lista de Gram negativas ($r=0.7768$). Correspondería con el candidato de m/z 453.1056 Da que eluye a 20.14.

Por último, destacan algunos compuestos de menor correlación que se han obtenido únicamente como candidatos frente a bacterias Gram positivas. El

²³⁰ Xie, Y.-Y. *et al.*, 2012. Comparative evaluation of cultivars of *Chrysanthemum morifolium* flowers by HPLC-DAD-ESI/MS analysis and antiallergic assay. *Journal of Agricultural and Food Chemistry*, 60(51), pp.12574–12583.



candidato de 433.0763 Da corresponde con la quercetina glicósido ($r=0.7766$), el candidato de 405.1180 Da corresponde con el tetrahidroxiestilbeno glucósido ($r=0.7713$), el candidato de 457.1169 Da es la daidzina 6"-O-acetato ($r=0.7671$) y el candidato de 615.0965 Da es la quercetina glicósido galato ($r=0.7504$).





BLOQUE III

Caracterización e identificación de compuestos anticancerígenos en muestras de aceite de oliva

BLOQUE 3. CARACTERIZACIÓN E IDENTIFICACIÓN DE COMPUESTOS ANTICANCERÍGENOS EN MUESTRAS DE ACEITE DE OLIVA.

Este último bloque de la memoria se presenta en lengua inglesa, si bien se incluye previamente un breve resumen del mismo en español.

Las muestras utilizadas para este estudio fueron aceites de oliva virgen extra de distintas zonas geográficas de España y distintas variedades de aceituna, incluyendo además aceites comerciales. A partir de las fuentes seleccionadas, se prepararon los extractos polifenólicos mediante extracción en fase sólida para ser posteriormente analizados en cuanto a su composición química y su bioactividad. Concretamente, los extractos se analizaron mediante HPLC-ESI-TOF-MS.

En este bloque, se presentan primeramente los resultados de la caracterización de los extractos, utilizando los datos tratados con el *clustering* de líneas y en segundo lugar, los resultados del estudio de correlación con la bioactividad.

En la caracterización realizada a partir del método de *clustering* de líneas se obtuvo una lista de *clusters* asociados a posibles compuestos, de entre los cuales se consiguió detectar los compuestos identificados en estudios de caracterización anteriores realizados con el procedimiento convencional. En el apéndice 3 se presentan las tablas con los parámetros utilizados en el *clustering* de líneas y la lista de compuestos obtenida, donde se han marcado en gris los compuestos identificados en la bibliografía.

Para el estudio de correlación con la bioactividad, se tomaron los datos de actividad anticancerígena de los extractos polifenólicos de aceite frente a la línea celular de cáncer de mama JIMT-1. La correlación con la bioactividad se realizó partiendo de los datos tratados mediante dos tipos de *clustering*: *clustering* rectangular y *clustering* de líneas.



En el estudio de correlación con los datos de *clustering* rectangular, se identificaron los candidatos a compuestos bioactivos y, en base a dichos resultados, se estimó un modelo de predicción de la bioactividad. Para ello, las muestras se separaron en dos grupos. El primer grupo de "entrenamiento", compuesto por 14 muestras, se utilizó para la identificación de candidatos y la estimación del modelo predictivo. El otro grupo de 8 muestras de "evaluación" se utilizó para la comparación de la bioactividad experimental y la predicha utilizando el modelo estimado anteriormente. El análisis de correlación se realizó entre la bioactividad y el área cromatográfica de cada intervalo de tiempo de retención y de m/z (tomando ventanas de 1 Da y 1 s), tomando los datos de las muestras del grupo de "entrenamiento". De este modo, se identificaron los candidatos bioactivos, con los que se estimó el modelo para predecir la bioactividad en muestras nuevas. Concretamente, se identificaron tres isómeros de la oleuropeína aglicona como candidatos a ser los responsables de la actividad anticancerígena de los extractos de aceite de oliva. Con el grupo de muestras de "evaluación", se validó el modelo predictivo, el cual proporcionó una estimación precisa de la bioactividad de las muestras nuevas.

En el estudio de correlación con los datos de *clustering* de líneas, los *clusters* asociados a los compuestos candidatos se obtienen de forma directa y sólo hay que identificarlos mediante su huella espectrométrica. Con este método se obtuvieron los mismos compuestos candidatos a ser los responsables de la bioactividad que los obtenidos con el *clustering* rectangular. De esta forma, se verifica la potencialidad de la herramienta de correlación para la identificación de compuestos bioactivos, así como la rapidez y robustez del *clustering* de líneas.



Characterization and identification of potentially anticarcinogenic compounds in olive oil samples.

Olive oil is one of the food matrices extensively studied in our research group, including studies of both characterization of bioactive compounds (mainly, phenolic compounds) and evaluation of different bioactive properties. In collaboration with the Catalan Institute of Oncology (ICO), many studies have been developed to demonstrate the *in vitro* anticancer activity of phenolic compounds from olive oil²³¹⁻²³⁴.

Previous studies of olive oil extracts have been based on fractionation and subsequent bioactivity measurement of the fractions collected in order to discover the bioactive compounds. In the study of García-Villalba *et al.* (2009), an olive oil extract was prepared by combining two varieties (Picual, rich in secoiríoides, and Arbequina, rich in lignans). A method of liquid chromatography (HPLC) and capillary electrophoresis (CE) was performed to separate and characterize each isolated fraction. After characterization, cell viability assays with SKBR3 breast cancer cell line was used to assess the bioactivity of the different fractions of the chromatographic profile. Figure 109 shows the images concerning the most bioactive fractions, corresponding with the fractions whose major compounds are secoiríoides²³⁵.

²³¹ Menéndez, J.A. *et al.*, 2008. Analyzing effects of extra-virgin olive oil polyphenols on breast cancer-associated fatty acid synthase protein expression using reverse-phase protein microarrays. *International Journal of Molecular Medicine*, 22, pp.433–439.

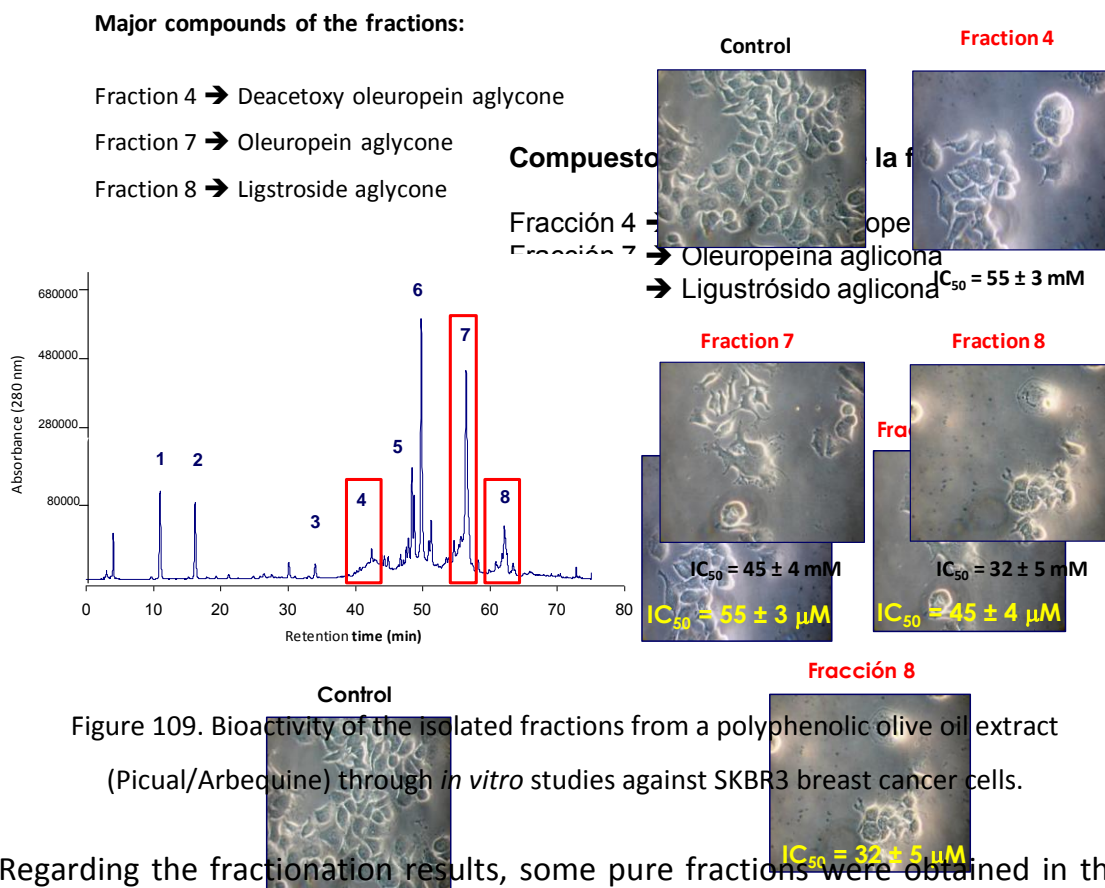
²³² Carrasco-Pancorbo, A. *et al.*, 2005. Evaluation of the antioxidant capacity of individual phenolic compounds in virgin olive oil. *Journal of Agricultural and Food Chemistry*, 53(23), pp.8918–8925.

²³³ Menéndez, J.A. *et al.*, 2007. Olive oil's bitter principle reverses acquired autoresistance to trastuzumab (Herceptin) in HER2-overexpressing breast cancer cells. *BMC Cancer*, 7, pp.80–99.

²³⁴ Carrasco Pancorbo, A. *et al.*, 2004. Sensitive determination of phenolic acids in extra-virgin olive oil by capillary zone electrophoresis. *Journal of Agricultural and Food Chemistry*, 52(22), pp.6687–6693.

²³⁵ García-Villalba, R. *et al.*, 2009. A 2-D-HPLC-CE platform coupled to ESI-TOF-MS to characterize the phenolic fraction in olive oil. *Electrophoresis*, 30(15), pp.2688–2701.





Regarding the fractionation results, some pure fractions were obtained in this study, although other fractions contained besides the main compound, other compounds in a lower concentration. Because of the difficulty of obtaining pure fractions, further studies are needed to find the compound responsible for the bioactivity.

Therefore, the main objective of this third section of the PhD thesis was to contribute to the characterization and identification of bioactive compounds from olive oil samples. For this, the data processing tools based on *clustering* and correlation were evaluated with a collection of HPLC-MS data from olive oil extracts, previously analyzed and characterized in our research group. The polyphenolic extracts were obtained from 22 extra virgin olive oil samples by using solid-phase extraction (SPE). The phenolic composition of the extracts was analyzed by using RRLC-ESI-TOF-MS, and their bioactivity was also evaluated by using cell viability assays with the JIMT-1 breast cancer cell line.



Firstly, this section presents the results of applying the *clustering* method to the MS data of the olive oil extracts in order to compare them with the conventional characterization. Secondly, the correlation strategy was evaluated by using the MS data from the extracts and the cytotoxicity index obtained in the bioassays with JIMT-1 breast cancer cells (see Figure 110).

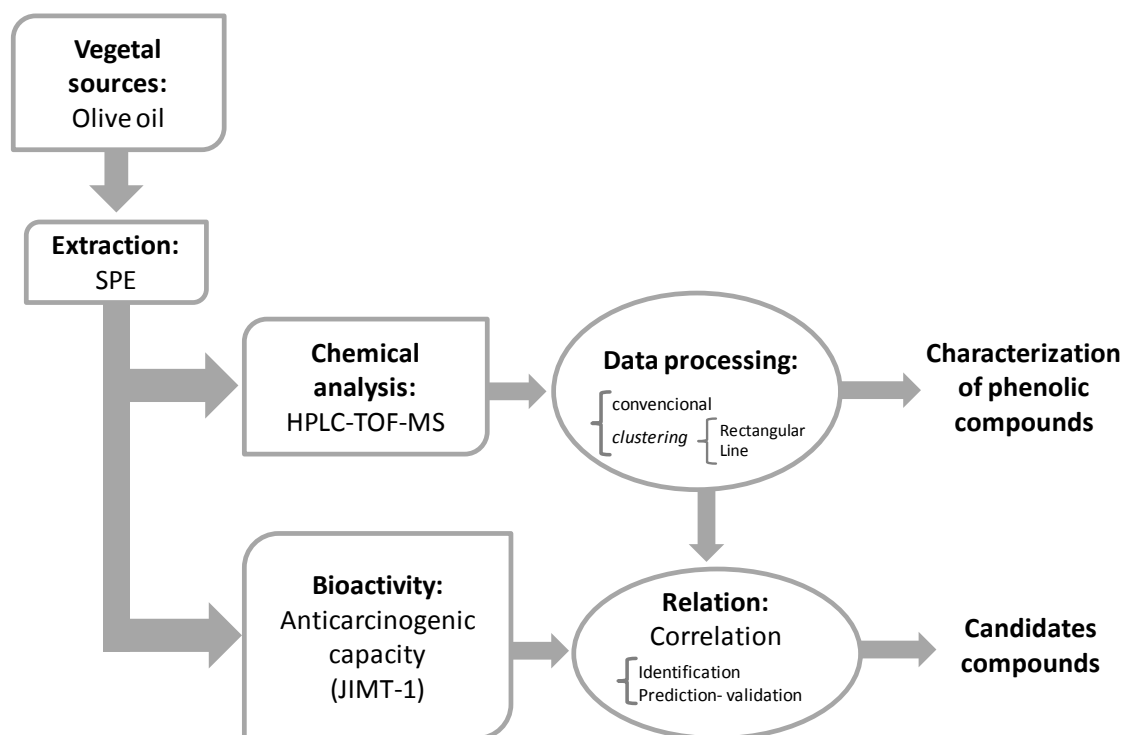


Figure 110. Outline studies in section 3.

3.1. Characterization of phenolic compounds in olive oil samples through the line-*clustering* method.

The first step of this study was the characterization with the list of possible compounds provided with the line-*clustering* method. The main goal of this characterization was to verify that all the compounds identified in previous studies are found in the list of compounds.

Figure 111 shows the 3D spectrogram of one of the extracts of the collection, while Figure 112 shows the BPC and TIC chromatograms and the low resolution



spectrogram (1s, 1Da) of the same extract, but using a colour scale for representing the intensity as a function of the retention time (t) and the m/z ratio. This graphical representation shows more information about the mass spectrum than the 3D one. The spectrogram can be expanded in order to study interesting portions of the LC-S data, as can be seen in Figure 113. This portion contains the main compounds of the spectrogram corresponding to the m/z range of [365 Da, 430 Da] and the retention time range of [12 min, 24 min].

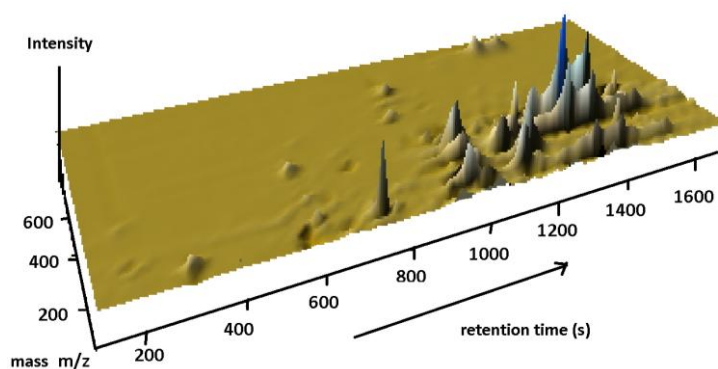


Figure 111. 3D spectrogram representation of one of the olive oil phenolic extract of the study.

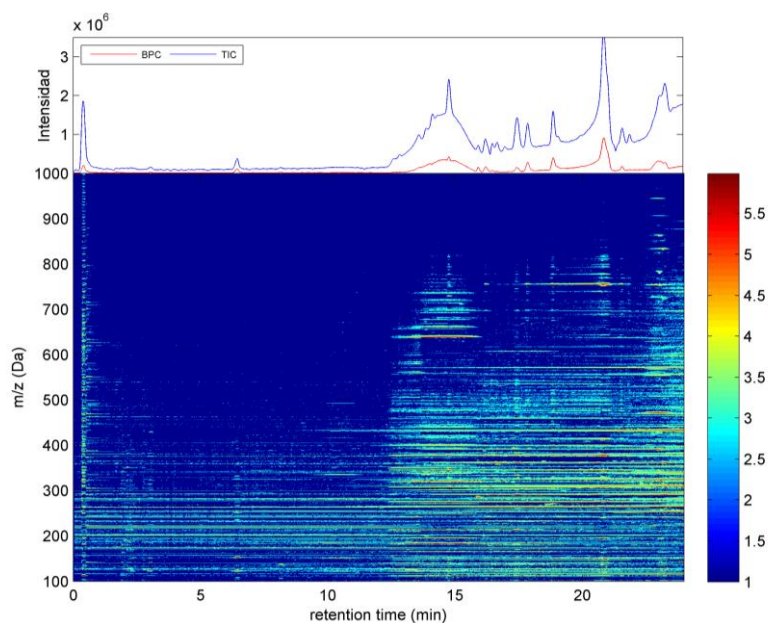


Figure 112. BPC and TIC chromatograms and low resolution spectrogram (1 s, 1 Da) of one of the olive oil phenolic extracts of the study. For each interval of elution time and m/z ratio, the intensity is represented according to the colour scale.



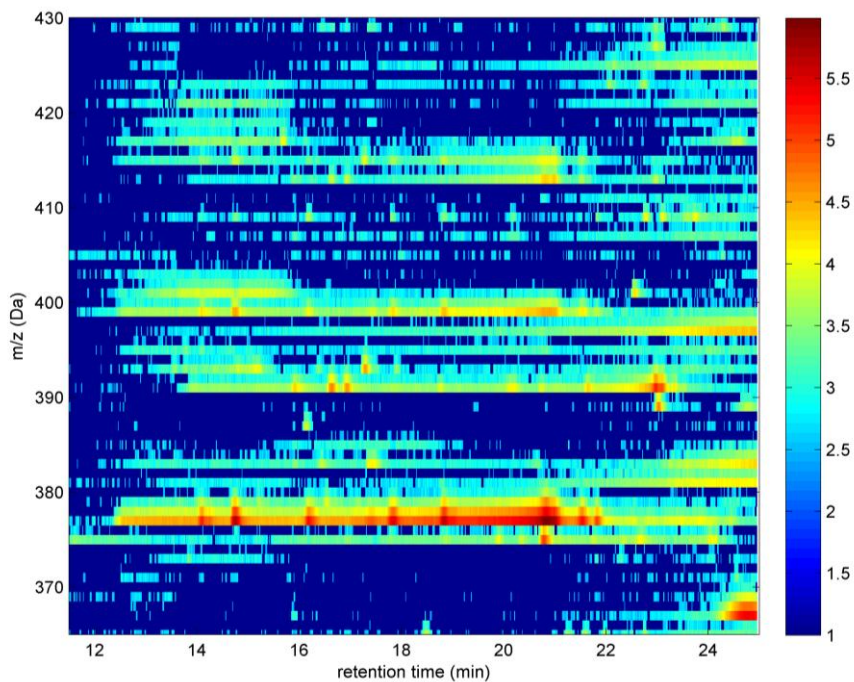


Figure 113. Detail of the zone corresponding to m/z 377 Da in the spectrogram of the figure 112.

In order to validate the *line-clustering* method, the data from the HPLC-MS analyses of the extracts was processed with the preprocessing and the *line-clustering* methods. As a result, a list of *clusters* associated to possible compounds was obtained. These compounds are represented in Figure 114.



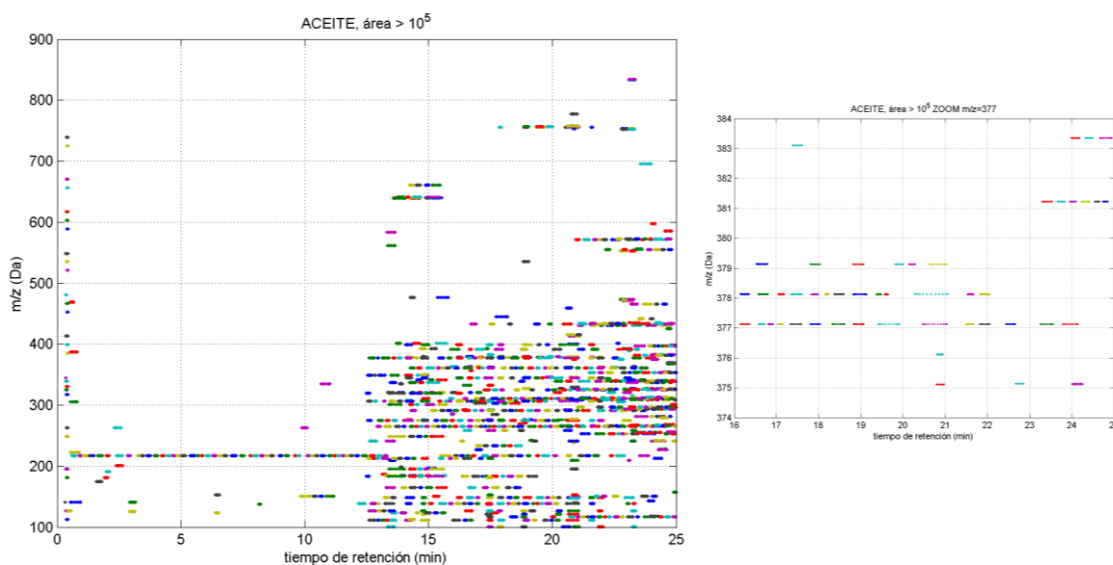


Figure 114. MS data obtained by applying the *line-clustering* method with one of the olive oil phenolic extract of the study. Threshold area $> 10^5$. A detail at right.

For one of the olive oil extracts analyzed, Figure 115 shows the base-peak chromatogram (blue) and all the peaks depicting the families of lines found through the *line-clustering* (red) when an area threshold of 1×10^5 is established. At the beginning of the chromatographic profile (until 13 min), few major compounds can be found and the baseline corresponding to the compounds eluting along the overall analysis time can also be seen. When these peaks eluting continuously are removed, the peaks of chromatogram in Figure 116 are obtained. As shown in this figure, there are many minor peaks that remain hidden from 13 min of the chromatogram. By applying the *clustering* method, 433 peaks corresponding to the representative peak of the families were detected.



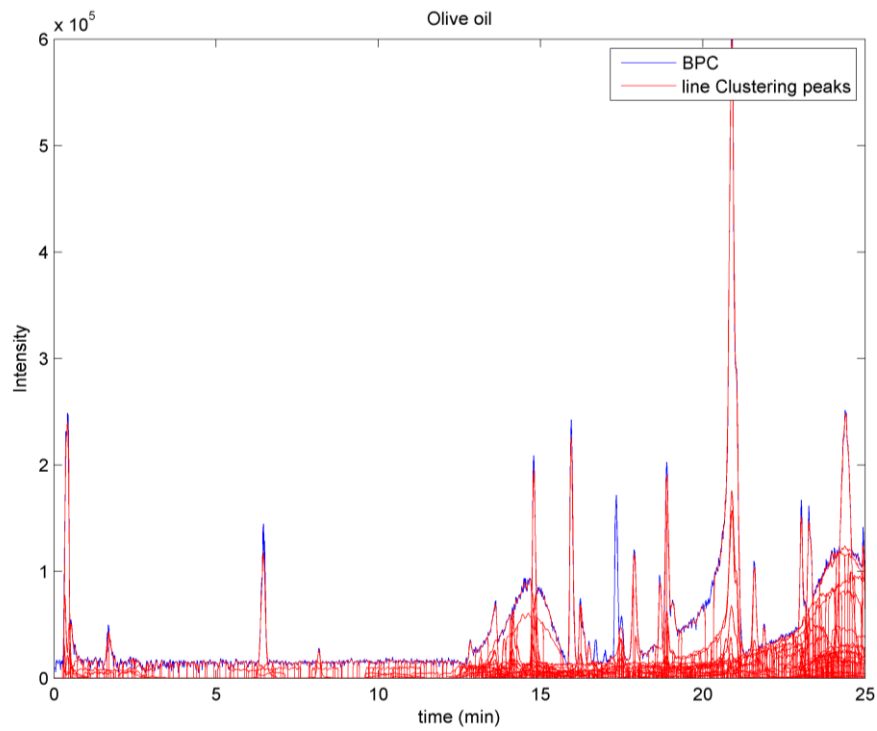


Figure 115. Base-peak chromatogram (blue) of one olive oil sample and all the peaks depicting the families of lines (red) found through the line-clustering, area threshold of 1×10^5 .

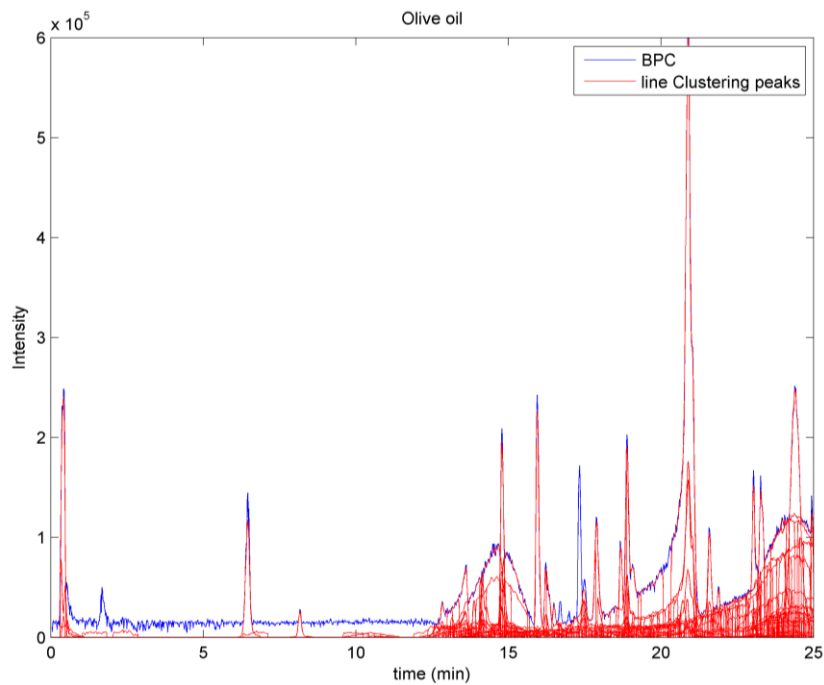


Figure 116. Base-peak chromatogram (blue) of the olive oil sample and all the peaks depicting the families of lines (red) found through the line-clustering, area threshold of 1×10^5 and removal of peaks eluting continuously.



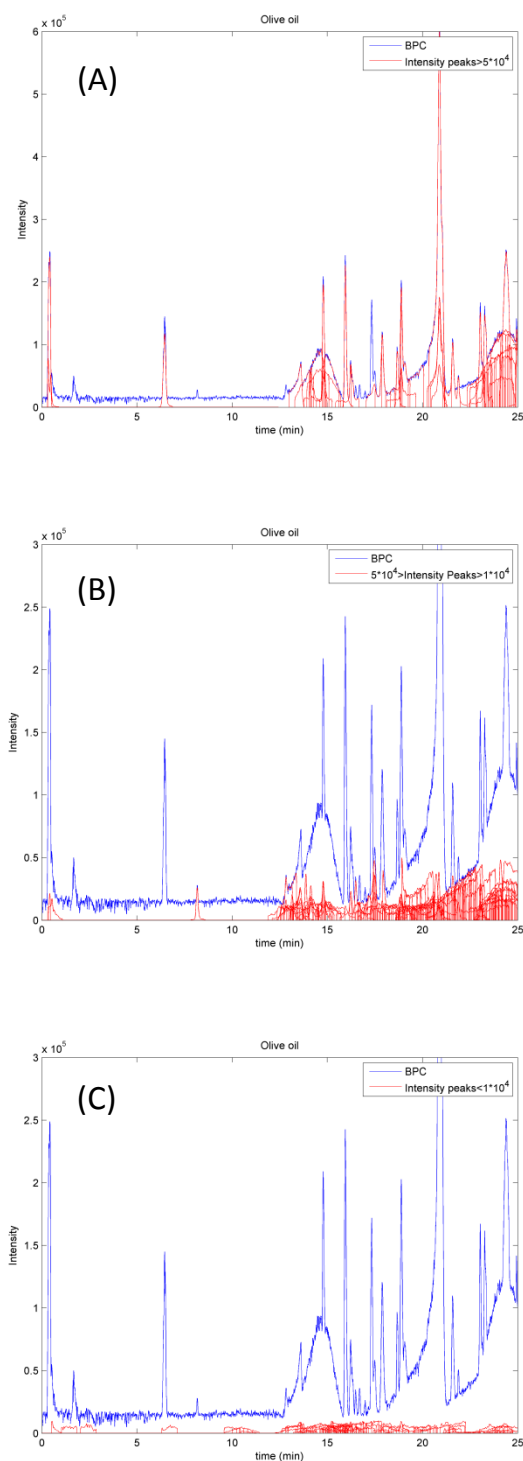


Figure 117. Base-peak chromatogram (blue) of the olive oil sample and all the peaks depicting the families of lines (red) found through the line-clustering, area threshold of 1×10^5 . (A) intensity peaks $> 10^5$ (54 peaks), (B) $10^5 > \text{Intensity} > 10^4$ (257 peaks), and intensity peaks $< 10^4$ (122 peaks).



Figure 117 shows three plots corresponding to the peaks whose intensity is greater than 10^5 (54 peaks), the peaks with intensity values between 10^5 and 10^4 (257 peaks), and the peaks whose intensity is lower than 10^4 (122 peaks).

As shown in these plots, the number of peaks detected by the *clustering* is much higher than using the conventional method due to the large amount of coeluting peaks that are hidden under the BPC. The next step is to identify every peak detected by *clustering*.

The compounds provided in the list of hundred of *clusters* were compared with the results of a previous work. The characterization results reported by Fu *et al.* (2009)²³⁶ can be found in the tables of the appendix 3. Among the 136 possible compounds generated by the *line-clustering*, all the compounds previously identified by using the conventional characterization method with these olive oil samples were detected.

3.2. Identification of bioactive compounds through the correlation method between the MS and the bioactivity data.

This correlation study is divided into three sections. In the first one, the correlation method used the HPLC-MS data previously processed with a *rectangular clustering* (1 s and 1 Da). The second section is based on the direct correlation between the bioactivity data and the list of *clusters* associated to compounds, previously provided by the *line-clustering* method. Finally, the third section includes the predictive potential and the validation of the correlation method.

²³⁶ Fu, S. *et al.*, 2009. Tentative characterization of novel phenolic compounds in extra virgin olive oils by rapid-resolution liquid chromatography coupled with mass spectrometry. *Journal of Agricultural and Food Chemistry*, 57(23), pp.11140–1117.



The correlation method requires from a collection of samples, a HPLC–MS analysis and a measurement of the bioactivity for each sample. By applying a correlation analysis between the bioactivity and the chromatographic area for each interval of retention time and m/z ratio, the peaks corresponding to candidates for bioactive compounds can be identified. Additionally, when peaks are identified, a model can be estimated to predict the bioactivity in new samples.

A high IC_{50} value corresponds to an extract in which the concentration of the bioactive compound is low. In order to assure a linear correlation with the concentration, a bioactivity parameter has been established by using a constant value in the numerator of the equation for the bioactivity data of all the samples. Therefore, the bioactivity data of the olive oil samples used for the correlation analysis is not the IC_{50} but a bioactivity parameter was previously defined according to:

$$y_n = \frac{0.1\%}{(IC_{50})_n}, \quad (\text{Equation 13})$$

where $(IC_{50})_n$ is the IC_{50} for sample n and y_n is the activity relative to an extract presenting an IC_{50} value of 0.1%.

Table 36 shows for each sample, the variety of EVOO, the IC_{50} parameter measured (v/v referred to the phenolic extract of 60 g of oil reconstituted in 1 mL of ethanol) and the bioactivity parameter. Samples were randomly assigned into two disjoint groups. One group with 14 samples (those identified with the label "training") was used to perform the correlation analysis, identify the candidates for the bioactive compound and develop the prediction model. The remaining 8 samples (identified with the label "evaluation") were used to



evaluate the identified candidates and to compare the measured bioactivity with the predicted bioactivity using new samples.

Table 36. Bioactivity of the polyphenolic olive oil extracts against the JIMT-1 breast cancer cell line. For each extract, the measured IC₅₀ and the bioactivity parameter are indicated.

Extract	IC ₅₀ (v/v)	Bioactivity	Group *
1	0.035%	2.86	Training
2	0.025%	4.00	Evaluation
3	0.038%	2.63	Training
4	0.030%	3.33	Training
5	0.026%	3.85	Evaluation
6	0.025%	4.00	Training
7	0.015%	6.67	Evaluation
8	0.026%	3.86	Evaluation
9	0.030%	3.33	Evaluation
10	0.022%	4.55	Training
11	0.028%	3.57	Evaluation
12	0.090%	1.11	Training
13	0.078%	1.28	Evaluation
14	0.070%	1.43	Training
15	0.037%	2.70	Training
16	0.032%	3.13	Training
17	0.066%	1.52	Training
18	0.056%	1.79	Training
19	0.025%	4.00	Training
20	0.035%	2.86	Evaluation
21	0.023%	4.35	Training
22	0.044%	2.27	Evaluation



3.2.1. Correlation with the rectangular *clustering*.

The application of the correlation method with the rectangular *clustering* is based on the HPLC–MS data and the measurements of bioactivity for a collection of samples (those selected for training). Figure 118 shows the base peak chromatogram and a representation of the chromatographic area $A_2(t_i, m_j)$ for sample 2 (one of the most active samples).

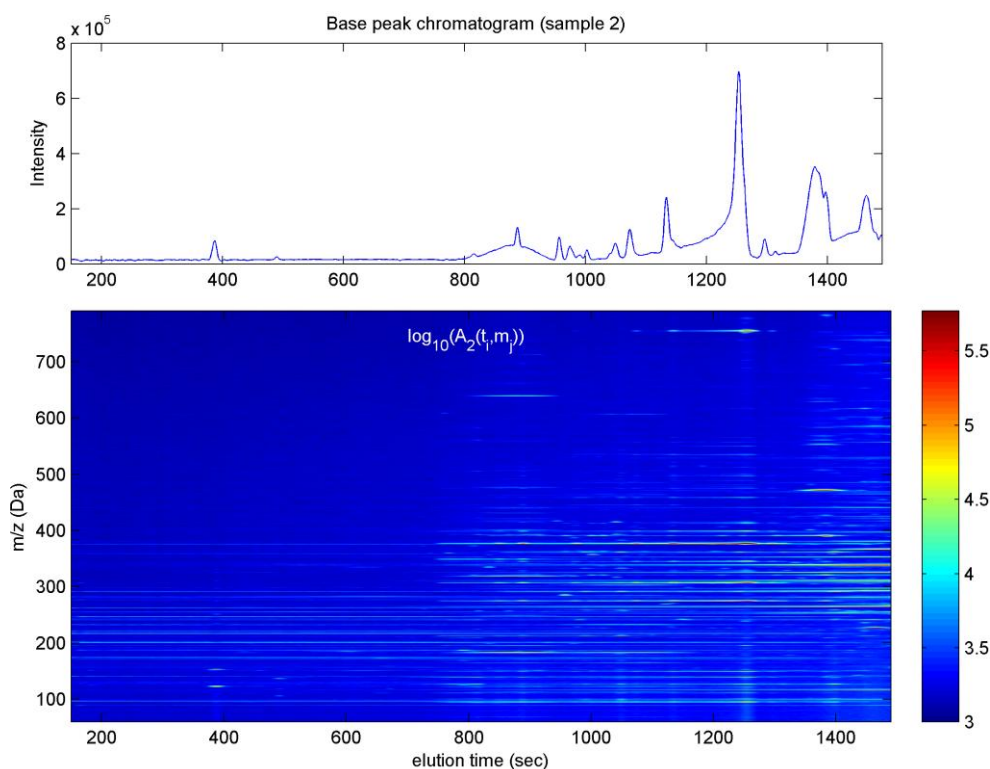


Figure 118. Description of sample 2 in experiment involving EVOO-PE. Top plot: base peak chromatogram. Bottom plot: representation of the area as a function of retention time and m/z. HPLC-MS data are represented after bucketing using $A_2(t_i, m_j)$ with resolutions of 1 s and 1 Da. Area represented in logarithmic scale according to the colour scale.



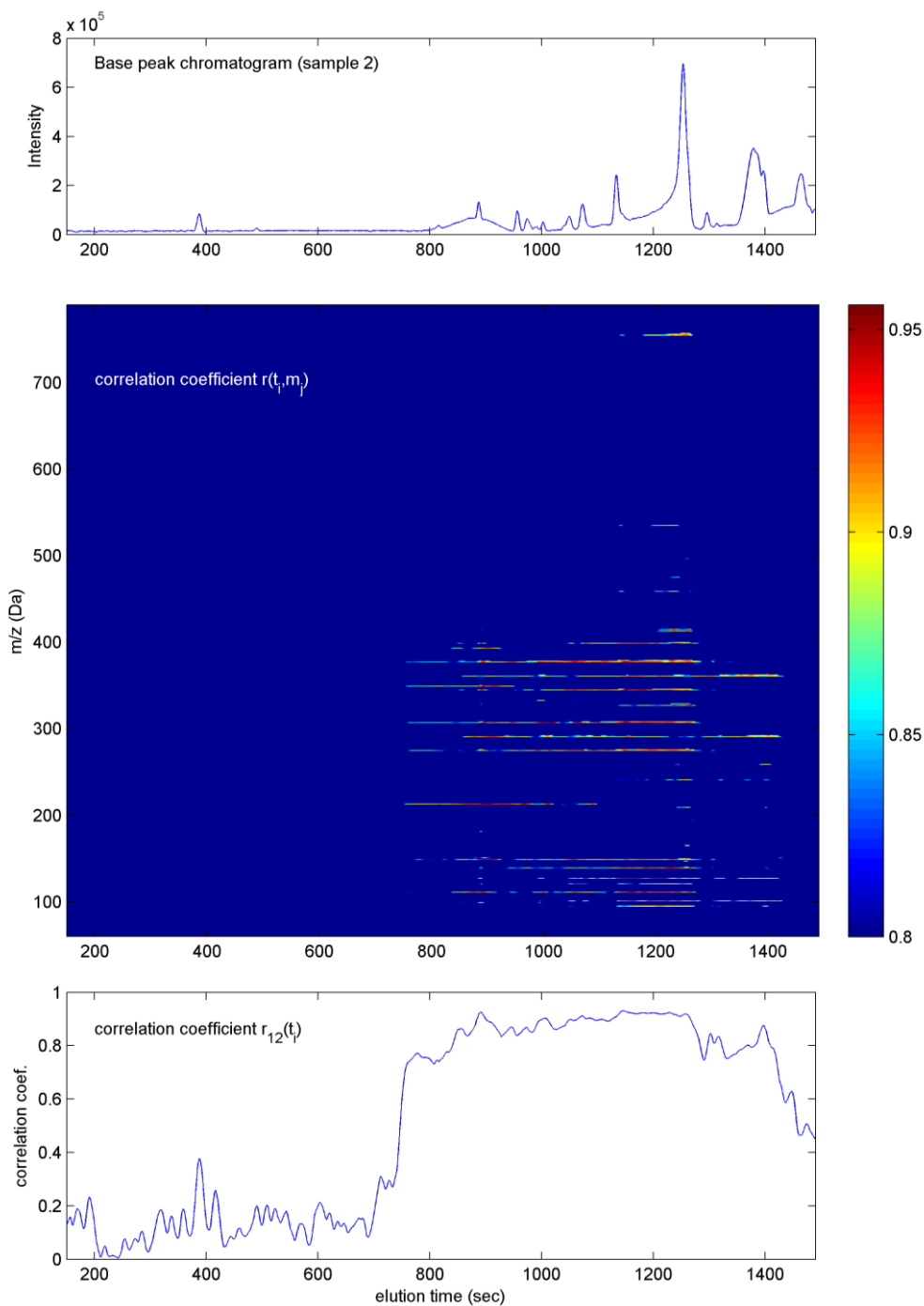


Figure 119. Correlation coefficient for the collection of samples in the training partition. The top plot contains the base peak chromatogram (for sample 2). In the middle, the correlation coefficient $r(t_i, m_j)$ between the area $A_n(t_i, m_j)$ and the bioactivity y_n is represented as a function of the elution time and the m/z ratio. The plot at the bottom represents the function $r_{12}(t_i)$ average of the best 12 correlation values of $r(t_i, m_j)$ at elution time t_i .



For the olive oil sample 2, Figure 119 shows the base peak chromatogram. It also shows the correlation coefficient $r(t_i, m_j)$ between the area $A_n(t_i, m_j)$ and the bioactivity y_n and the function $r_{12}(t_i)$ average of the best 12 correlation values of $r(t_i, m_j)$ at elution time t_i . Two important families of high correlation lines can be found in Figure 119. The first one corresponds to the interval of retention time [800 s, 1280 s] (the m/z value with high correlation and maximum intensity in this interval is $m_j = 377$ Da). The second one presents a retention time around 1400 s (high correlation and maximum intensity for $m_j = 361$ Da). There are more families, but with lower correlation (and also some spurious peaks) in the correlation plot.

The function $r_k(t_i)$ was calculated (and represented in Fig. 4, bottom plot) using Equation 12 with $K=12$. Again, the correlation reached high values in the interval [800 s, 1280 s] and around 1400 s, and $r_{12}(t_i)$ is smaller than 0.85 out of these regions. The identification of candidates was based on an analysis of the functions $r_k(t_i)$ and $r_k(t_i, m_j)$ at each retention time t_i . A candidate was proposed at t_i when $r_k(t_i)$ was greater than a threshold (which was set to 0.85) and the most abundant of the K spectral peaks involved in the estimation of $r_k(t_i)$ also had a correlation coefficient greater than this threshold. This most abundant peak provides the m/z ratio, the area and the correlation coefficient of possible candidates when the previous conditions are satisfied.

Figure 120 shows the m/z ratio, the chromatographic area and the correlation coefficient of those high correlation most abundant spectral peaks as a function of the retention time t_i . From this plot, several candidates for the active compound were identified: six peaks with m/z ratio $m_j = 377$ Da (eluting at 887, 976, 1047, 1074, 1137 and 1254 s), one peak with $m_j = 393$ Da eluting at 910 s, and 2 peaks with $m_j = 361$ Da (eluting at 993 and 1397 s). The raw high resolution HPLC–MS data (for retention time intervals and m/z ratios around the



peaks identified as candidates) provides the accurate m/z ratio of the candidates and the mass spectrum profile.

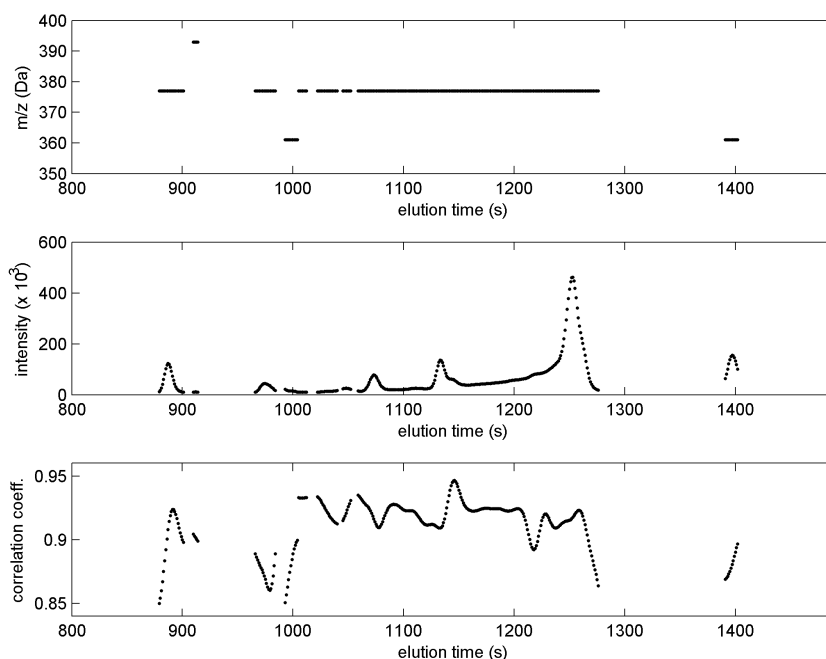


Figure 120. Identification of candidates for the bioactive compound. A point was generated in the plots when $r_{12}(t_i) > 0.85$ and the most abundant of the 12 selected spectral peaks also showed a correlation coefficient greater than 0.85. The plots represent the m/z ratio (top), the area (middle) and the correlation coefficient (bottom) of this most abundant spectral peak, as a function of elution time when the above conditions are satisfied.

Taking into account the accurate m/z ratios and retention times, the candidates with 377 Da are isomers of oleuropein aglycone ($C_{19}H_{22}O_8$, exact m/z : 377.12419121 Da; experimental m/z : between 377.1241 Da and 377.1254 Da). The candidates with 361 Da are isomers of ligstroside aglycone ($C_{19}H_{22}O_7$, exact m/z : 361.12910583 Da; experimental m/z : between 361.1294 Da and 361.1305 Da), and the candidate with 393 is 10-hydroxy oleuropein aglycone ($C_{19}H_{22}O_9$, exact m/z : 393.11910583 Da; experimental m/z : 393.1195 Da). In this analysis, the list of candidates could have included more peaks (by applying a smaller threshold). However, the correlation coefficient decreased for the subsequent



candidates, and therefore the bioactive compound is expected to be one of those identified with this figure.

3.2.2. Correlation with the *line-clustering*.

The correlation strategy was also evaluated in the experiment involving the same olive oil extracts, but using the analytical data previously processed with the *line-clustering*. The area of the *clusters* associated with possible compounds provided in the list was calculated for each *cluster* in every sample. The area values were correlated with the bioactivity data using the 14 samples (training) and the correlation coefficients were obtained as a function of the retention time and the m/z ratio. These values can be represented in a correlogram (Figure 121).

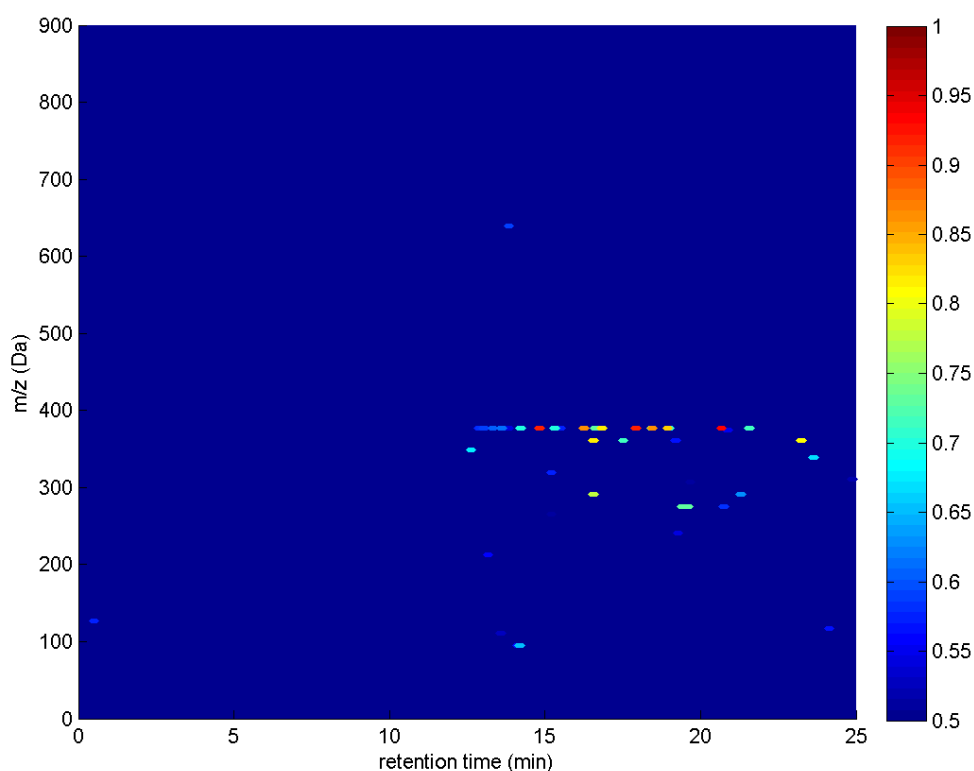


Figure 121. Correlation coefficient for the collection of 14 olive oil polyphenolic extracts (training). The correlation coefficient $r(t_i, m_j)$ between the area $A_n(t_i, m_j)$ and the bioactivity y_n is represented as a function of the retention time and the m/z ratio.



The red *clusters* corresponding to high correlation represent the possible candidates to bioactive compounds. These candidates are specified in Table 37.

Table 37. List of candidates compounds to be responsible for the bioactivity, by using the LC-MS data processed with the *line-clustering*.

#	r	m/z (Da)	t _R (min)
1	0.9351	377.1259	20.67
2	0.9173	377.1247	17.92
3	0.9169	377.1249	14.82
4	0.8625	377.1246	16.26
5	0.8613	377.1244	18.45
6	0.8547	299.0559	19.32
7	0.8302	377.1251	18.94
8	0.8182	361.1294	16.55
9	0.8171	377.1245	16.83
10	0.8103	361.1307	23.25
11	0.8055	415.1414	17.34

It was verified that the candidates responsible for bioactivity were the same as those obtained by correlating with the rectangular *clustering* data.

3.2.3. Model for prediction of the bioactivity.

Regarding the prediction of bioactivity of this correlation study, a prediction model is developed. The areas of the 9 chromatographic peaks selected as candidates were measured in the samples in the training partition. The chromatographic area \mathcal{A}_{nc} for sample n ($n = 1, \dots, 14$) and candidate c ($c = 1, \dots, 9$) was measured with a narrow window (± 10 mDa) around the m/z ratio of the candidates, and in the expected retention time interval. One prediction model was estimated for each candidate, by linear regression between the



chromatographic area \mathcal{A}_{nc} and the activity y_n . Each candidate c was evaluated by comparing the correlation coefficient between the area \mathcal{A}_{nc} and the activity y_n when the samples in the training partition or those in the evaluation partition are considered. For each candidate, Table 38 shows both correlation coefficients. Candidates 1, 2, 3, 7, 8 and 9 present a correlation coefficient in the evaluation partition that is significantly lower than that in the training partition. This is a strong argument to discard them as possible active compounds. On the other hand, candidates 4, 5 and 6 (that are three isomers of oleuropein aglycone) can be considered valid candidates for the active compound.

Table 38. Validation of the identified candidates for the bioactive compound. For each candidate c , the correlation coefficients between the chromatographic area and the bioactivity corresponding to the training and evaluation partition are compared. A valid candidate should present a high correlation coefficient in both partitions.

C	m/z (Da)	Retention time (s)	r (train)	r (eval)	Valid?
1	377.1254	887	0.922	0.554	No
2	377.1245	976	0.940	0.755	No
3	377.1241	1047	0.928	0.765	No
4	377.1245	1074	0.921	0.949	Yes
5	377.1248	1137	0.941	0.972	Yes
6	377.1252	1254	0.942	0.962	Yes
7	393.1195	910	0.899	0.226	No
8	361.1294	993	0.881	0.589	No
9	361.1305	1397	0.898	0.732	No

Figure 122 represents the activity as a function of the chromatographic area for candidate 6 (the isomer of oleuropein aglycone eluting at 1254 s). The regression line (providing the predicted activity) and the limits of the 95% confidence interval are also shown. The unfilled and filled circles represent the



samples in the training and evaluation groups, respectively. The difference between the measured activity and the regression line represent the error of the prediction model for each sample. As can be seen, the regression line provides a reasonable prediction for the samples in both the training and the evaluation groups.

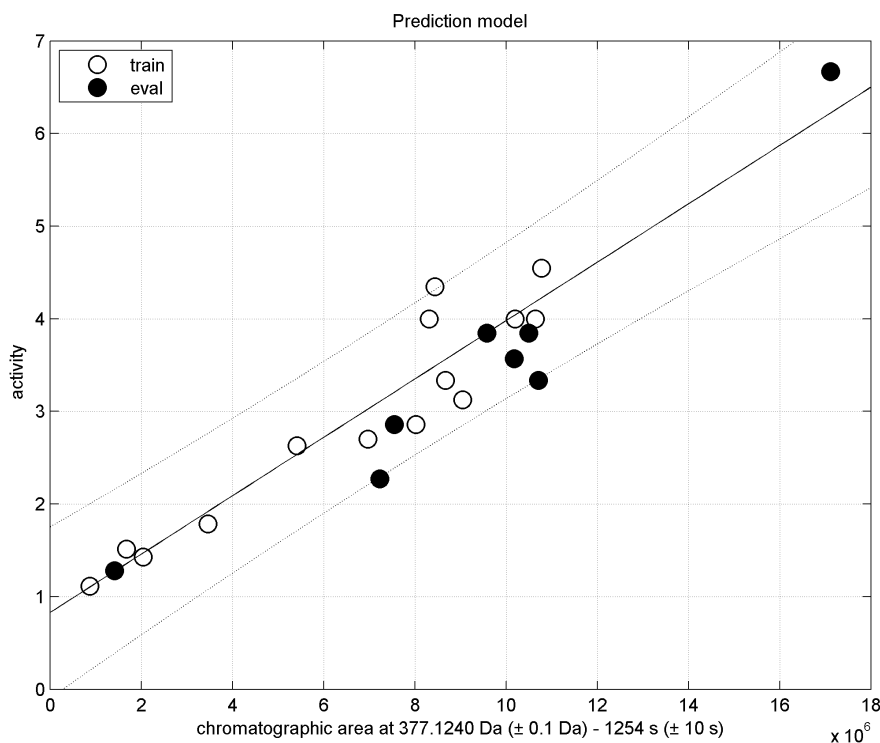


Figure 122. Bioactivity (cytotoxicity against JIMT-1 breast cancer) as a function of the chromatographic area \mathcal{A}_n for candidate 6 (m/z ratio: 377.1252 Da; elution time: 1254 s). Unfilled and filled circles represent the samples in the training and evaluation partitions, respectively. The solid and dotted lines are the regression line and the limits of the 95% confidence interval (estimated from the training partition).

The main results of the regression analysis between the chromatographic areas and the bioactivities for the valid candidates (4, 5 and 6) are shown in Figure 122. The accuracy of the prediction procedure can be evaluated with the root mean squared error (rmse) between the measured bioactivity and the predicted bioactivity. The rmse relative to the standard deviation of the bioactivity has



been included for both the training and the evaluation partitions. Similar errors are found for candidates 5 and 6, and these are smaller than for candidate 4.

Table 39. Results of the regression analysis for each valid candidate. For each candidate, the rmse relative to the standard deviation of the bioactivity (rel rmse) for the training and evaluation partitions, the intercept, the slope (and the corresponding standard errors) and the semi-width of the 95% confidence interval are shown.

	Candidate 4	Candidate 5	Candidate 6
m/z (Da)	377.1245	377.1248	377.1252
Retention time (s)	1074	1137	1254
Retention time (min)	17.9	18.9	20.9
Rel rmse (train)	37.5%	32.5%	32.4%
Rel rmse (eval)	76.4%	47.70%	44,20%
Intercept	11.412	0.94952	0.8307
SE (intercept)	0.25449	0.23406	0.24396
Slope	$2,136 \times 10^{-6}$	$1,1019 \times 10^{-6}$	$3,1501 \times 10^{-7}$
SE (slope)	$2,6095 \times 10^{-7}$	1.14×10^{-7}	$3,2444 \times 10^{-8}$
S.w. conf int 95 %	10.011	0.86604	0.86444

The table also includes, for each candidate, the values of the intercept, the slope, the corresponding standard errors and the semiwidth of the 95% confidence interval. The results lead to the conclusion that, with a high probability, one of these isomers of oleuropein aglycone is the bioactive compound, while the others are compounds with concentrations that are highly correlated with that of the bioactive one. The most probable candidates are those eluting at 1137 s and 1254 s. The predictions from both provide a similar accuracy, with a relative rmse in the prediction of bioactivity close to 45% for the evaluation partition and close to 30% for the training partition. The relatively high values of the correlation coefficients and the relatively low error provided by the prediction models suggest the utility of the methods to identify the candidates for the bioactive compound and to predict the bioactivity of new samples.





CONCLUSIONES

Conclusiones generales de la tesis:

- Se ha seleccionado una gran variedad de fuentes potenciales de compuestos bioactivos de origen vegetal, incluyendo frutas, vegetales, aceites y plantas aromáticas. A partir de las especies vegetales, utilizando diferentes sistemas de extracción y secado, se han obtenido amplias colecciones de extractos con las que realizar estudios analíticos y de bioactividad.
- Para analizar la composición de los extractos, se han utilizado diferentes técnicas separativas (GC, HPLC y UPLC) acopladas a espectrometría de masas con distintos detectores (Q, TOF y QTOF). Los datos obtenidos de dichos análisis se han procesado mediante el método convencional (usando el *software* del fabricante), proporcionando resultados de caracterización satisfactorios.
- Para evaluar la bioactividad de las colecciones de extractos, se han realizado bioensayos *in vitro*, incluyendo valoración de capacidad antioxidante, antimicrobiana y anticancerígena. Con ello, se han obtenido conjuntos de datos de bioactividad variable, poniendo de manifiesto el potencial biológico de muchos de los extractos estudiados.
- Se han desarrollado novedosas herramientas avanzadas de tratamiento de datos de cromatografía acoplada a espectrometría de masas, incluyendo reducción de ruido y *clustering*. Se ha demostrado que con la utilización de dichas herramientas, se consigue compactar los datos de espectrometría de masas, reducir considerablemente el tiempo de procesamiento y obtener de manera automática una lista de posibles compuestos, incluyendo la detección de aquellos no identificados en la caracterización convencional debido al fenómeno de coelución.



- Se ha desarrollado una herramienta de procesamiento que, mediante la correlación entre los datos de espectrometría de masas y los datos de bioactividad de los extractos vegetales, es capaz de proporcionar una lista de compuestos candidatos a ser responsables de la bioactividad.
- Las herramientas de *clustering* y correlación se han evaluado en tres estudios concretos, implicando tres tipos diferentes de especies vegetales y datos de bioactividad procedentes de tres tipos de ensayos biológicos: extractos de pimiento y capacidad antioxidante, extractos de *Cistus* y actividad antimicrobiana y, por último, extractos de aceite de oliva y actividad anticancerígena. En todos los casos, han funcionado de manera satisfactoria, demostrando la utilidad de las herramientas propuestas para la caracterización de colecciones de extractos vegetales y la identificación de compuestos bioactivos.

Conclusiones específicas de las herramientas desarrolladas de tratamiento y procesamiento de datos de espectrometría de masas:

- El algoritmo de reducción de ruido presentado en esta tesis es un procedimiento simple y efectivo para eliminar datos del espectro de masas que, con alta probabilidad, corresponden a ruido instrumental. Los resultados experimentales muestran que el algoritmo resulta efectivo, proporcionando espectros de masas más limpios (con menor ruido y por tanto, más adecuados para un posterior procesamiento, ya sea manual, supervisado o automático) y con un volumen de datos significativamente menor de manera que se facilita el almacenamiento y procesamiento posterior, sin riesgo pérdida de datos relevantes.
- La novedad y la aportación más importante del tratamiento propuesto de datos de espectrometría de masas es el *preclustering* y la obtención de

una lista final de *clusters* o posibles compuestos que facilita considerablemente la labor de detectar compuestos, incluyendo compuestos mayoritarios, minoritarios, e incluso aquellos picos ocultos bajo otros de mayor intensidad que no son detectables con el método convencional debido a la coelución de los compuestos.

- El algoritmo de *clustering* desarrollado se ha aplicado a diferentes casos, demostrando ser un tratamiento rápido, automático y versátil de los datos de espectrometría de masas y poniendo de manifiesto su gran utilidad para la caracterización de extractos vegetales. Esta herramienta de tratamiento los datos tiene una utilidad potencial para conjuntos de muchas muestras, tareas como comparación de espectros de masas, metabolómica, alineamiento automático de espectros, etc.
- La herramienta de correlación de datos de espectrometría de masas y datos de bioactividad de colecciones de extractos constituye un método útil para la identificación de candidatos potencialmente responsables de la bioactividad. Con respecto al método convencional, este método requiere únicamente un dato de bioactividad por extracto, prescindiendo por tanto, de los tediosos procesos de fraccionamiento de extractos y la posterior valoración de la bioactividad de cada fracción. Además de las ventajas de tipo económico y ahorro de tiempo que proporciona este método, destaca su versatilidad y aplicabilidad a todo tipo de productos naturales y propiedades bioactivas.

Conclusiones específicas del bloque 1:

- Se han seleccionado cuarenta y cuatro variedades de frutas y vegetales procedentes del sur de España, siendo recolectadas a diferentes tiempos. Se ha evaluado la capacidad antioxidante de las mismas, mediante la



utilización de tres métodos diferentes (TEAC, FRAP y ORAC), proporcionando una idea de la capacidad antioxidante que aporta el consumo de dichas variedades a lo largo del año. Algunas de las muestras analizadas han demostrado ser importantes fuentes de compuestos antioxidantes. Entre las frutas que obtuvieron mejores resultados, destacan el limón, la uva roja y la granada, mientras que, entre los vegetales, los mejores resultados se obtuvieron con las variedades de pimiento rojo y amarillo y el ajo blanco.

- En base a los resultados de capacidad antioxidante, se seleccionaron tres variedades de pimiento (*Capsicum annum* L.) para la realización de un estudio más exhaustivo de sus constituyentes antioxidantes. Se analizaron mediante HPLC-DAD-ESI-TOF-MS, demostrando la complementariedad de la información proporcionada por la espectroscopía UV-vis y la espectrometría de masas con analizador de tiempo de vuelo para la identificación de compuestos fenólicos y otros compuestos polares. La caracterización se realizó con el método convencional, lográndose identificar un total de 45 compuestos, de los cuales 34 son descritos en esta matriz alimentaria por primera vez. Por ello, la información obtenida de la caracterización puede ser de gran utilidad para la actualización de las tablas de composición de alimentos españolas.
- Se han tratado los análisis de pimiento mediante las herramientas de *clustering* de líneas, obteniendo una lista de posibles compuestos ordenados por área. Se ha comprobado que, entre los 164 compuestos primeros de la lista obtenida, se encuentran los compuestos previamente identificados con el método convencional.



- A partir de los datos preprocesados con el *clustering* de líneas, se ha evaluado una variante de la herramienta de correlación en este caso en el que se dispone de pocas muestras, para poder identificar los candidatos a compuestos responsables de la capacidad antioxidante. Los candidatos obtenidos fueron muy similares al correlacionar los datos TEAC y FRAP con los datos analíticos y distintos de la correlación realizada con los resultados ORAC, posiblemente debido al mecanismo de acción en el que se basan los métodos. Entre los candidatos proporcionados para los métodos TEAC y FRAP, destacan los ácidos hidroxicinámicos, flavonoles y los ácidos orgánicos, mientras que para el método ORAC la correlación fue menor, obteniendo un coeficiente de correlación relativamente alto para el ácido quínico.

Conclusiones específicas del bloque 2:

- Se ha estudiado el perfil de compuestos volátiles de diferentes especies españolas del género *Cistus*. La microextracción en fase sólida seguida del análisis por cromatografía de gases acoplada a espectrometría de masas (SPME-GC-MS) ha demostrado ser un sistema de extracción apropiado para la extracción y el análisis en continuo de volátiles. Se identificó un total de 112 compuestos volátiles, la gran mayoría de naturaleza terpénica y de los cuales, 28 son descritos por primera vez en *Cistus*.
- Se ha estudiado el perfil polifenólico de extractos de *Cistus* obtenidos mediante diversos métodos de extracción, secado y purificación. Se analizaron mediante dos plataformas analíticas distintas, HPLC-TOF-MS y UPLC-QTOF-MS, poniendo de manifiesto la ventaja que implica el sistema de detección QTOF frente al TOF, dado que ofrece una mayor información estructural. Mediante la caracterización convencional, se ha

obtenido un total de 58 compuestos identificados, 18 de los cuales se describen por primera vez en *Cistus*.

- Se ha evaluado la capacidad antimicrobiana de los extractos polifenólicos de *Cistus* utilizando un modelo de bacteria Gram negativa (*E. coli*) y un modelo de bacteria Gram positiva (*S. aureus*). Los resultados de estos ensayos de bioactividad pusieron de manifiesto el potencial de los extractos de *Cistus* como ingredientes antimicrobianos en aplicaciones tales como cosmética e higiene, destacando la capacidad antimicrobiana frente a *S. aureus* de los tres extractos acuosos de *C. salviifolius* obtenidos mediante spray drying.
- Se ha aplicado la herramienta *clustering* de líneas a una amplia colección de extractos polifenólicos de *Cistus*. En primer lugar, se seleccionaron tres de los extractos más bioactivos para realizar un estudio comparativo entre los resultados de la caracterización convencional y la caracterización realizada con la lista de compuestos obtenida en el *clustering* de líneas. Se ha comprobado la presencia de todos los compuestos identificados por el procedimiento convencional, entre los 184 primeros de la lista obtenida de cientos de *clusters* asociados a posibles compuestos. Gracias a la aplicación del *clustering* de líneas, se ha conseguido identificar 17 compuestos nuevos, 10 de los cuales se encontraban ocultos bajo otros compuestos de mayor intensidad.
- Para el estudio de correlación, en el que se utilizó una colección de 26 extractos polifenólicos de *Cistus*, se encontraron 7 compuestos cuya concentración está asociada estadísticamente a la actividad frente a bacterias Gram positivas y 8 compuestos cuya concentración está asociada estadísticamente a la actividad antimicrobiana con el modelo de bacteria Gram negativa. Entre los posibles compuestos responsables de

la actividad antimicrobiana, los elagitaninos se proponen como candidatos frente a ambas cepas, *E. coli* y *S. aureus*, obteniendo coeficientes de correlación más altos en el caso de *E. coli*. Frente a *S. aureus*, se obtuvieron candidatos principalmente de la familia de los flavonoides.

Conclusiones específicas del bloque 3:

- Se ha utilizado una colección de extractos polifenólicos de aceite de oliva virgen extra, previamente estudiada en nuestro centro de investigación. Los datos procedentes del análisis de cromatografía acoplada a espectrometría de masas de dichos extractos se trataron con el *clustering* de líneas, realizando la caracterización a partir de dichos datos tratados. De esta forma, se comprobó que se obtienen los compuestos caracterizados en estudios anteriores, verificando la fiabilidad de la herramienta desarrollada.
- Para el estudio de correlación con la bioactividad, se han tomado los datos de capacidad anticancerígena de dichos extractos frente a la línea celular de cáncer de mama JIMT-1. En este caso, la herramienta de correlación propuesta, además de identificar candidatos a compuestos bioactivos, demostró su potencialidad predictiva, de manera que, al incluir una muestra nueva al estudio, es capaz de predecir a partir de los datos analíticos, la bioactividad que presentará. Se ha aplicado la herramienta de correlación utilizando los datos tratados mediante dos métodos de *clustering* distintos: *clustering* de líneas y *clustering* rectangular. Con ambos métodos, se obtuvieron resultados similares en cuanto a los candidatos a compuestos con actividad anticancerígena, si bien el *clustering* de líneas ha demostrado ser un método más rápido y robusto.

- Los candidatos más prometedores de la bioactividad frente a la línea celular de cáncer de mama JIMT-1 fueron tres isómeros de la oleuropeína aglicona. Esta información debe contrastarse con el correspondiente proceso de fraccionamiento y evaluación de la bioactividad de la fracción que contenga los compuestos candidatos resultantes del estudio, quedando este trabajo como perspectiva futura.





CONCLUSIONS

General conclusions of the PhD thesis:

- Potential plant sources of bioactive compounds have been selected including fruits, vegetables, oils and herbs. Extensive collections of extracts were obtained from the plant species by using different extraction and drying systems in order to perform analytical and bioactivity studies with the extracts.
- Different separation techniques (GC, HPLC and UPLC) coupled to mass spectrometry with different mass analyzers (Q, QTOF and TOF) have been used to analyze the composition of the extracts. The data obtained from these analyzes were processed by the conventional method (commercial software) and characterization provided satisfactory results.
- *In vitro* bioassays, including antioxidant, antimicrobial and anti-cancer capacity assays have been performed in order to assess the bioactivity of the extracts. The biological potential of many of the extracts was shown, and datasets of variable bioactivity were obtained.
- New advanced tools for data processing of chromatography coupled to mass spectrometry data have been developed, including noise reduction and *clustering*. With the use of such tools, the mass spectrometry data is compacted, the processing time is greatly reduced and a list of possible compounds is automatically obtained, including those unidentified compounds by the conventional characterization because of the phenomenon of coelution.
- A new processing tool for the identification of bioactive compounds has been developed. It is based on the correlation between mass spectrometry data and bioactivity data of plant extracts and it has shown



to be capable of providing a list of candidates to be responsible for the bioactivity.

- The *clustering* and correlation tools have been evaluated in three specific studies involving three different types of plants and bioactivity data from three types of biological assays: pepper extracts and antioxidant capacity, *Cistus* extracts and antimicrobial activity, and olive oil extracts and anticancer activity. In all cases, results were satisfactory and the usefulness of the tools proposed for the characterization of plant extracts and the identification of bioactive compounds was demonstrated.

Specific conclusions of the proposed tools for processing mass spectrometry data:

- The noise reduction algorithm presented in this thesis is a simple and effective method for removing mass spectral data, which correspond to instrumental noise with high probability. Experimental results showed that the algorithm provides mass spectra with less noise and therefore, more suitable for further processing (either manually, monitored or automatically) and a significantly smaller amount of data without loss of significant data. This facilitates the storage and subsequent processing.
- The novelty and the most important contribution of the proposed tools is the *preclustering* and the list of *clusters* or possible compounds obtained. This tool has demonstrated to be useful for detecting compounds, including major and minor compounds, and even those peaks hidden under other larger peaks, which would not be detectable with the conventional method.
- The *clustering* algorithm has been applied to different cases. It has been proven to be a fast, versatile, and automatic method for the



characterization of plant extracts. This data processing tool is mainly useful for the processing of sets of many samples, comparison of mass spectra, metabolomics, automatic alignment of spectra, etc.

- The correlation tool is useful for identifying candidates potentially responsible for the bioactivity from collections of mass spectrometry and bioactivity data of extracts. With respect to the conventional method, the correlation tool requires only one bioactivity data from each extract, regardless tedious fractionation processes of extracts and subsequent measurement of the bioactivity of the fractions. In addition to the economic and time advantages provided by this method, its versatility and applicability to all kinds of natural products and bioactive properties is highlighted.

Specific conclusions of Subsection 1 of Results and Discussion:

- Forty-four cultivars of fruits and vegetables from southern Spain have been selected, being harvested at different times. The antioxidant capacity of these samples have been evaluated using three different methods (TEAC, FRAP, and ORAC) to give an idea of the antioxidant capacity provided by the intake of these cultivars year round. Some of the samples have demonstrated to be important sources of antioxidant compounds. The fruits that showed the highest antioxidant capacity were lemon, red grapes and pomegranate, while among the vegetables, the best results were obtained with cultivars of red and yellow pepper and white garlic.
- Three varieties of pepper (*Capsicum annuum* L.) showing high antioxidant capacity were selected to carry out a comprehensive study of its antioxidant constituents. The polar fraction of these pepper samples has



been analyzed by HPLC–DAD-ESI-TOF-MS, showing the complementarity of the information provided by UV-Vis spectroscopy and mass spectrometry with time-of-flight analyzer for identifying phenolic compounds and other polar compounds. The conventional characterization method was used and a total of 45 compounds were tentatively identified, 34 of which being reported in this food matrix for the first time. Therefore, the information obtained from the characterization could be useful to update the data of the Spanish composition databases.

- The *line-clustering* was used for processing the HPLC-MS analyzes of the pepper extracts. A list of possible compounds sorted by area was obtained. The compounds previously identified by the conventional method have been found between the first 164 compounds of the list.
- From the data preprocessed with the *line-clustering*, a variant of the correlation tool has been evaluated in order to identify candidate compounds responsible for the antioxidant capacity. Candidates obtained by using the TEAC and FRAP data were different from candidates obtained with ORAC results, possibly due to the mechanism of action on which the methods are based. The correlation for the ORAC method was lower than the other antioxidant capacity methods. Among the candidates provided for the FRAP and TEAC data, feruloyl glucoside is highlighted, whereas quinic acid was the compound of higher correlation for the ORAC method.

Specific conclusions of Subsection 2 of Results and Discussion:

- The profile of volatile compounds from different Spanish species of the genus *Cistus* has been studied. The solid-phase microextraction followed



by gas chromatography coupled to mass spectrometry analysis (SPME-GC-MS) has proven to be an appropriate extraction system for the extraction and analysis of volatile continuously. A total of 112 volatile compounds was tentatively identified, most of them being terpenes and 28 of which being reported for the first time in *Cistus*.

- The phenolic profile of *Cistus* extracts obtained by different methods of extraction, drying and purification was studied. The use of two different analytical techniques, HPLC-TOF-MS and UPLC-QTOF-MS demonstrated the inherent advantage of QTOF detection system versus TOF, because it offers greater structural information. The conventional characterization allowed identifying a total of 58 compounds, 18 of which are reported in *Cistus* for the first time.
- The antimicrobial capacity of polyphenolic *Cistus* extracts has been assessed by using a Gram negative bacteria model (*E. coli*) and a Gram positive model (*S. aureus*). The results of these bioactivity assays showed the potential of *Cistus* extracts as antimicrobial ingredients in applications such as cosmetics and hygiene. The antimicrobial activity against *S. aureus* of the aqueous extracts of *C. salviifolius* obtained by spray drying is highlighted.
- The line-*clustering* tool has been applied to a wide collection of polyphenolic extracts from *Cistus*. First, three of the most bioactive extracts were selected to compare the results of the conventional characterization and the characterization performed with the list of compounds obtained in the line-*clustering*. The occurrence of all the compounds identified by the conventional process, among the first 184 compounds obtained in the list of *clusters* associated with possible compounds has been verified. By applying the line-*clustering*, it has been



able to identify 17 new compounds, 10 of which were hidden under other compounds of greater intensity.

- For the correlation study, a collection of 26 polyphenolic extracts of *Cistus* was used. Seven compounds whose concentration is statistically associated with the activity against Gram positive were found as well as eight compounds whose concentration is statistically linked to the activity against Gram negative bacteria. Among the compounds potentially responsible for the antimicrobial activity, ellagitannins were proposed as candidates against both *E. coli* and *S. aureus* strains with higher correlation coefficients in the case of *E. coli*. Against *S. aureus*, candidates were mainly flavonoids.

Specific conclusions of Subsection 2 of Results and Discussion:

- A collection of polyphenolic extracts from extra virgin olive oil, previously studied in our research centre was used. Data from the analyses of chromatography coupled to mass spectrometry of these extracts were processed with the *line-clustering* in order to perform the characterization. All the compounds identified in previous studies were obtained; therefore, the reliability of this tool was verified.
- For the correlation study with bioactivity, anticarcinogenic capacity data of such extracts against JIMT-1 breast cancer cell line was used. In this case, the utility of the proposed correlation tool was demonstrated not only for identifying bioactive candidates, but also for predicting the bioactivity of new samples from their analytical data. The correlation tool has been applied by using the data processed with two different *clustering* methods: *line-clustering* and *rectangular clustering*. With both methods, similar results were obtained in terms of the bioactive



candidates, although the *line-clustering* has proven to be a faster and more robust method.

- Three isomers of oleuropein aglycone were found to be possibly responsible for the antitumor activity against breast cancer of type JIMT-1. This information should be contrasted with the corresponding process of fractionation and the evaluation of the bioactivity of the fraction containing the candidates resulting from the study.



ABCDEFGHIJ
KLMNOPQR
STUVWXYZ

0123456789

%!?'()*=-*\$\$€@

ABREVIATURAS

- AAPH:** 2,2'-azobis-(2-aminopropano)-dihidrocloruro
- ABTS:** ácido 2,2'-azinobis (3-etilbenzotiazolín)-6-sulfónico
- ADN:** ácido desoxirribonucleico
- AUC:** área bajo la curva
- BPC:** cromatograma de pico base
- CoA:** coenzima A
- D.I.C.:** descompresión instantánea controlada
- DAD:** detector de batería de diodos
- dm/z:** desviación estándar en masa
- DPPH:** 2,2-difenil-1-picrilhidracilo
- dt_R:** desviación estándar en tiempo
- EI:** impacto electrónico
- EIC:** cromatograma de ion extraído
- EM:** electromultiplicador
- ESI:** ionización por electrospray
- EVOO:** aceite de oliva virgen extra (*extra virgin olive oil*)
- FD:** liofilización (freeze drying)
- FRAP:** ferric ion reducing antioxidant power
- FWHM:** anchura de pico a mitad de su altura máxima
- GC:** cromatografía de gases
- HAT:** transferencia de átomos de hidrógeno (*hydrogen atom transfer*)
- HPLC:** cromatografía líquida de alta resolución
- HS:** espacio de cabeza (*headspace*)
- IC₅₀:** concentración requerida para alcanzar la del efecto inhibitor máximo
- IK:** índice de Kovats
- INT:** p-iodonitrotetrazolium
- IPP:** isopentenil difosfato



ISCID: disociación inducida por colisión en la fuente (*in-source collision-induced dissociation*)

LC: cromatografía líquida

LDA: análisis discriminante lineal

LLE: extracción líquido-líquido

m/z: relación masa/carga

MAE: extracción asistida por microondas

MAGRAMA: Ministerio de Agricultura, Alimentación y Medio Ambiente

MEP: metileritritol fosfato

MIC₅₀: concentración mínima inhibitoria requerida para inhibir el crecimiento de organismos en un 50%

MS: espectrometría de masas

MTT: bromuro de 3-(4,5-dimetiltiazol-2-il)-2,5-difeniltetrazolium

ORAC: oxigen radical absorbance capacity

PCA: análisis de componentes principales

PLE: extracción con fluidos presurizados

Q: analizador de masas cuadrupolo

Q-TOF: analizador de masas cuadrupolo-tiempo de vuelo

RE: receptor de estrógenos

RF: radiofrecuencia

rmse: error cuadrático medio (*root mean squared error*)

RRLC: cromatografía líquida de resolución rápida

SDY: spray drying

SE: desviación estándar

SET: transferencia de un electrón desapareado (*single electron transfer*)

SFE: extracción con fluidos supercríticos

SLE: extracción sólido-líquido

SPE: extracción en fase sólida



SPME: microextracción en fase sólida

TBARS: sustancias reactivas al ácido tiobarbitúricothiobarbituric acid reactive substances

TEAC: trolox equivalent antioxidant capacity

TIC: cromatograma de iones totales

TOF: analizador de masas de tiempo de vuelo

TPTZ: 4,6-tripiridil-s-triazina

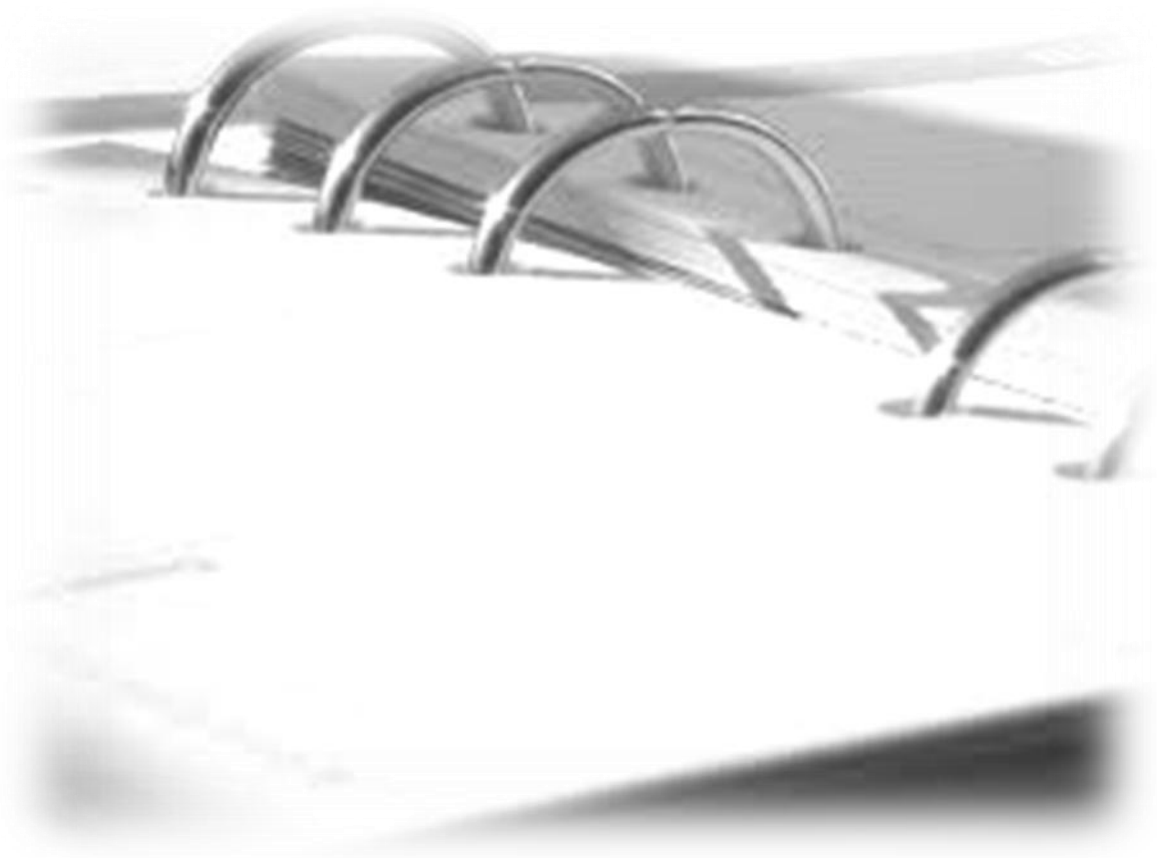
t_R: tiempo de retención

UAE: extracción asistida por ultrasonidos

UPLC: cromatografía líquida de ultra alta resolución

UV-Vis: ultravioleta-visible





ÍNDICE FIGURAS Y TABLAS

ÍNDICE DE TABLAS

Tabla 1. Extractos polifenólicos de aceite de oliva virgen extra de diferentes variedades de aceituna, diferentes aéreas geográficas de España. 117

Tabla 2. Extractos polifenólicos de diferentes especies de *Cistus* obtenidos mediante distintos métodos de extracción y secado. 122

Tabla 3. Esquema de las plataformas analíticas utilizadas para el análisis de compuestos fenólicos en la presente tesis: cromatografía líquida y espectrometría de masas. 141

Tabla 4. Efecto del algoritmo de reducción de ruido sobre el tamaño de un fichero binario (.bin) de un extracto vegetal de pimienta. El fichero binario se ha obtenido a partir de un fichero .baf del LC-MS que ocupa 553.29 MB..... 184

Tabla 5. Ejemplo de fichero de resultados estadísticos para la zona de la muestra de pimienta estudiada, comprendida en el intervalo de tiempo [1200 s, 2100 s] y de masa [446 Da, 450 Da]. 202

Tabla 6. Compuestos calculados mediante el *clustering* de líneas y el procesado posterior para encontrar familias de especies iónicas: ión principal y sus isótopos, correspondientes a la zona de estudio de la Figura 59 donde se ha filtrado en área > 10⁵ 207

Tabla 7. Resultado de aplicar el algoritmo de identificación de compuestos a través de las familias de líneas o huellas espectrométricas a la lista de *clusters* obtenida con el *clustering* de líneas a una muestra simple con 5 estándares. Se ha utilizado un valor de umbral_área_huella= 15.000 cuentas. Se muestra los *clusters* eliminados, y los compuestos ordenados por área y por tiempo de retención. Para cada compuesto se especifica: la relación m/z, y su desviación estándar (en Da), el tiempo de retención y su desviación estándar (en minutos y segundos), el área (en Kcuentas= 1000 cuentas). 214

Tabla 8. Compuestos obtenidos y 10 de las líneas de mayor área de su huella espectrométrica. Se especifica la masa (Da), el tiempo de retención (min), el área (Kcuentas) y el porcentaje relativo de área de cada una de las líneas que componen la huella espectrométrica respecto del ión de mayor área. En gris, el ión principal de las huellas espectrométricas. 215

Tabla 9. Comparación entre el resultado de *clustering* de líneas y los compuestos reales de la muestra. Se indican los nombres de los compuestos fenólicos, la fórmula molecular de su ión, su masa exacta calculada con la fórmula molecular, y los valores de m/z y t_R experimentales obtenidos mediante el *clustering* de líneas. 216

Tabla 10. Parámetros aplicados en el *clustering* de líneas. 219

Tabla 11. Compuestos fenólicos de las 30 muestras: se indica el código identificador, el nombre, la fórmula molecular y la masa exacta. 228

Tabla 12. Concentración de los 5 compuestos fenólicos (c1, c2, c3, c4 y c5) en cada una de las 30 muestras. Las dos últimas columnas representan los valores de absorbancia a 525nm de cada muestra y si la muestra se ha usado como entrenamiento o evaluación. Las últimas líneas muestran el coeficiente de correlación r entre la absorbancia y la concentración de los diferentes compuestos. Se ha calculado la



correlación con las muestras usadas para entrenamiento ($r(\text{entrenam.})$), para evaluación ($r(\text{evalua.})$) y para todas las muestras ($r(\text{todas})$).230

Tabla 13. Compuestos candidatos obtenidos mediante la correlación entre el área de los *clusters* del *clustering* rectangular y la actividad de las 20 muestras de entrenamiento del experimento.234

Tabla 14. Lista de candidatos a compuestos activos que eluyen en $t_r= 6.25$ min. Se indican las masas m/z experimentales y las áreas de los 12 picos espectrales con mayor coeficiente de correlación. Los picos están ordenados por intensidad.235

Tabla 15. Candidatos a ser responsables de la actividad, obtenidos mediante la correlación entre el área de los compuestos generados directamente mediante el *clustering* de líneas y la actividad de las 20 muestras de entrenamiento del experimento.237

Tabla 16. Resultados del análisis de regresión entre el área cromatográfica y la actividad para los 3 candidatos. Se comparan los coeficientes de correlación r obtenidos para las muestras de entrenamiento y evaluación. Para cada candidato se presenta, la desviación estándar relativa de la actividad (rel. rmse), la intersección con el eje y la pendiente (y su correspondiente desviación estándar) y la semi-anchura del intervalo de confianza del 95%.240

Tabla 17. Capacidad antioxidante (métodos TEAC, FRAP y ORAC) por 100 g (peso seco) y % de humedad para las muestras de frutas.250

Tabla 18. Capacidad antioxidante (métodos TEAC, FRAP y ORAC) por 100 g (peso seco) y % de humedad para las muestras de vegetales.254

Tabla 19. Compuestos identificados en los extractos de pimiento analizados mediante HPLC-ESI-TOF-MS: A (pimiento verde), B (pimiento amarillo) y C (pimiento rojo).275

Tabla 20. Compuestos desconocidos en los tres extractos de pimiento.278

Tabla 21. Datos de los ficheros de las muestras de pimiento.289

Tabla 22. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de pimiento.289

Tabla 23. Parámetros utilizados para el *clustering* de líneas de todas las muestras de pimiento analizadas.290

Tabla 24. Parte del listado de compuestos resultados del *clustering* de líneas donde aparecen los tres compuestos del ejemplo.293

Tabla 25. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método TEAC de las muestras de pimiento.300

Tabla 26. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método FRAP de las muestras de pimiento.301

Tabla 27. Compuestos candidatos a ser responsables de la capacidad antioxidante obtenida mediante el método ORAC de las muestras de pimiento.302

Tabla 28. Compuestos identificados en las cuatro especies de *Cistus* analizadas mediante SPME-GC-MS.317



Tabla 29. Capacidad antimicrobiana de los extractos de *Cistus* frente a *E. coli* (modelo de bacteria Gram negativa) y *S. aureus* (modelo de bacteria Gram positiva), expresada como concentración mínima inhibitoria 50 (MIC₅₀) en µg/mL.....334

Tabla 30. Compuestos identificados en los extractos de *C. salviifolius* analizados mediante RRLC-ESI-TOF-MS.....341

Tabla 31. Datos de los ficheros de las muestras de *Cistus* analizadas para la caracterización.....348

Tabla 32. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de *Cistus*. ...348

Tabla 33. Parámetros utilizados para el preprocesamiento de los ficheros de las muestras de *Cistus*. ...349

Tabla 34. Caracterización de los extractos de *Cistus salviifolius* S2, S2a y S4 mediante el tratamiento de los datos con el *clustering* de líneas.357

Tabla 35. Listas de compuestos candidatos a ser responsables de la capacidad antimicrobiana de los 26 extractos de *Cistus* frente a bacterias Gram negativas y Gram positivas.....369

Table 36. Bioactivity of the polyphenolic olive oil extracts against the JIMT-1 breast cancer cell line. For each extract, the measured IC₅₀ and the bioactivity parameter are indicated.389

Table 37. Validation of the identified candidates for the bioactive compound. For each candidate *c*, the correlation coefficients between the chromatographic area and the bioactivity corresponding to the training and evaluation partition are compared. A valid candidate should present a high correlation coefficient in both partitions.395

Table 38. Results of the regression analysis for each valid candidate. For each candidate, the rmse relative to the standard deviation of the bioactivity (rel rmse) for the training and evaluation partitions, the intercept, the slope (and the corresponding standard errors) and the semi-width of the 95% confidence interval are shown.397

Table 39. List of candidates compounds to be responsible for the bioactivity, by using the LC-MS data processed with the *line-clustering*.395



ÍNDICE DE FIGURAS

Figura 1. Extracto vegetal..... 51

Figura 2. Clasificación y algunas estructuras de terpenos..... 54

Figura 3. Clasificación y estructuras de los compuestos fenólicos. 57

Figura 4. Clasificación y estructura de los flavonoides. 58

Figura 5. Causas del estrés oxidativo, situaciones de sobreproducción de radicales libres. 61

Figura 6. Estructura del ácido 2,2'-azinobis (3-etilbenzotiazolín)-6-sulfónico (ABTS). 63

Figura 7. Reacción de reducción del TPTZ producida en el método FRAP. 64

Figura 8. Reacción de oxidación de la fluoresceína por el AAPH que se produce en el método ORAC. 65

Figura 9. Tinción de Gram (izquierda) y esquema simplificado de la pared celular bacteriana (derecha). Modelo de bacteria Gram negativa (superior) y Gram positiva (inferior.)..... 66

Figura 10. Métodos de extracción de compuestos bioactivos. 72

Figura 11. Sistema de microextracción en fase sólida de espacio en cabeza (HS-SPME)..... 77

Figura 12. Esquema de las partes de un cromatógrafo de gases. 80

Figura 13. Esquema básico de un equipo de HPLC junto a uno comercial. 83

Figura 14. Esquema de las partes de un espectrómetro de masas y su acoplamiento con cromatografía líquida o de gases. 85

Figura 15. Esquema del sistema de ionización por impacto electrónico..... 86

Figura 16. Esquema del proceso de formación del electrospray. 88

Figura 17. Esquema de un analizador de masas cuadrupolo (Q). 90

Figura 18. Esquema detallado de las partes de un espectrómetro de masas TOF..... 91

Figura 19. Esquema detallado de las partes de un espectrómetro de masas QTOF. 93

Figura 20. Esquema de los estudios realizados en la tesis. 109

Figura 21. Esquema general de la metodología llevada a cabo. 110

Figura 22. Esquema de todas las variedades de frutas y vegetales que conforman el estudio de capacidad antioxidante y la temporalidad de las mismas. 113

Figura 23. Colección de 26 extractos polifenólicos de *Cistus* para el estudio de correlación. 123

Figura 24. Equipo de cromatografía líquida de resolución rápida (RRLC) acoplado a espectrómetro de masas con analizador de tiempo de vuelo (TOF-MS) mediante ionización por electrospray (ESI). 126

Figura 25. Equipo de UPLC Acquity Waters y microTOF Q-II de Bruker Daltonik. 130

Figura 26. Lector de microplacas Synergy™ Mx BioTek y microplaca de 96 pocillos para valorar la capacidad antioxidante. 134

Figura 27. Esquema del funcionamiento de un reflectrón de un espectrómetro de masas de tiempo de vuelo, TOF..... 147

Figura 28. Multiplicador de electrones o electromultiplicador 148



Figura 29. Arriba: ejemplo de una señal espectro de suma (para 6 *scans* sucesivos), donde se representa en el eje X la *m/z* en un intervalo de masa (*m/z*) [375 Da, 381 Da] y en el eje Y la intensidad registrada por el detector para 6 *scans* con $t_R = 1251 \text{ s} - 1256 \text{ s}$, para una muestra de extracto fenólico de aceite de oliva. Abajo: espectros de línea que calcula el espectrómetro a partir de los espectros de suma anteriores. La intensidad está en escala logarítmica..... 151

Figura 30. Fragmento de un fichero en formato XML conteniendo los datos de espectrometría de masas medidos en el laboratorio. 152

Figura 31. Ejemplo de datos de espectrometría de masas obtenidos mediante HPLC (fase inversa) acoplada a MS para un extracto polifenólico de aceite de oliva. 153

Figura 32. Izquierda: intensidad frente a la masa del ión *m/z* (Da); en azul un espectro de masas (para 1250.39 s); en verde el siguiente (1251.39 s); en rojo el siguiente (1252.40 s). A la derecha, un detalle para un intervalo de masas pequeño entre 375 y 381 Da. La última gráfica representa la intensidad en escala logarítmica. 154

Figura 33. Señales proporcionadas en un análisis LC-MS. A la derecha: cuatro espectros de masas (*m/z*-I) en diferentes colores para cuatro tiempos de retención determinados (t_R). 155

Figura 34. Representación típica de datos LC-MS mediante cromatogramas para un extracto de aceite de oliva. (A) Cromatograma de ión total o TIC: corriente total en función del tiempo de retención para el rango de masas $m/z = 50$ a 1200 Da; (B) Cromatograma de pico base, BPC: corriente correspondiente al ión más abundante en función de t_R ; (C) Cromatograma de ión extraído, EIC: corriente correspondiente a un rango de masas (ión extraído) $m/z = 377.125 \pm 0.010$ Da en función del t_R 156

Figura 35. Izquierda: Espectrograma de baja resolución (1 Da, 1 s) para una muestra de un extracto fenólico de aceite de oliva. Derecha: detalle en una zona del espectrograma en el intervalo de tiempo [11 min, 23 min] y de masa [370 Da, 384 Da]. 158

Figura 36. Izquierda: Espectrograma de alta resolución (50 mDa, 1 s) para una muestra de un extracto fenólico de aceite de oliva. Derecha: detalle en una zona del espectrograma en el intervalo de tiempo [11 min, 23 min] y de masa [370 Da, 384 Da]. 158

Figura 37. Arriba: cromatograma Base Peak de una muestra simple formada por un solo compuesto, ácido caféico; abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra. 160

Figura 38. Detalle sobre el espectrograma de baja resolución de la Figura 37 de una muestra simple de 1 compuesto. 161

Figura 39. Arriba: cromatograma Base Peak de una muestra simple formada por dos compuestos, delfinidina-3-O-sambubiósido y ácido caféico. Abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra. 162

Figura 40. Detalle de la Figura 39 para el intervalo de tiempo [5 min, 7.5 min] y de masa de [580 Da, 710 Da]. Arriba: cromatograma BPC de la delfinidina-3-O-sambubiósido. Abajo, espectrograma de baja resolución (1 s, 1 Da). Se ha marcado el tiempo de retención $t_R = 6.25$ min correspondiente al máximo del pico cromatográfico. 163



Figura 41. Izquierda: espectros de línea $I(m/z)$ para el tiempo de retención $t_R = 6.25$ min de la Figura 40. Derecha: detalle del espectro de línea de la izquierda, para el ión principal de masa $m/z=595$ Da y sus isótopos $m/z=596, 597, 598$ Da. La intensidad está en escala logarítmica. 164

Figura 42. Arriba: cromatograma Base Peak de una muestra simple formada por seis compuestos; abajo: espectrograma de baja resolución (1 s, 1 Da) para la muestra simple..... 165

Figura 43. Arriba: cromatogramas BPC (en rojo) y TIC (en verde) del análisis LC-MS de un extracto fenólico complejo de pimiento. Abajo: espectrograma de baja resolución de dicho extracto vegetal. El cuadro marcado en rojo corresponde a un detalle que se muestra en la figura siguiente..... 169

Figura 44. Arriba: cromatogramas; BPC (en rojo) de una muestra vegetal compleja de un extracto vegetal de pimiento y cromatograma EIC (en verde) para una masa $m/z=597.1\pm 0.1$ Da. Abajo: espectrograma de baja resolución de la misma muestra vegetal. 170

Figura 45. Arriba: cromatograma EIC para $m/z=447.1\pm 0.1$ Da. Abajo: representación de la masa m/z en función del tiempo de retención, donde se aprecia que m/z exacta queda bien definida en la zona del pico donde tiene mucha intensidad. 171

Figura 46. Esquema seguido para el tratamiento de datos LC-MS desarrollado: métodos de *clustering* rectangular y *clustering* de líneas..... 175

Figura 47. Histograma de la intensidad de una muestra de un extracto vegetal de pimiento. 179

Figura 48. Cromatogramas EIC para $m/z= 449\pm 0.5$ Da de un extracto vegetal de pimiento, aplicando el algoritmo de reducción de ruido con distintos umbrales de intensidad, 100, 500, 1000 cuentas. 180

Figura 49. Espectrogramas de alta resolución (1 s, 10 mDa) para la masa $m/z=449$ Da de un extracto vegetal de pimiento analizado por LC-MS. Se ha aplicado el algoritmo de reducción de ruido con distintos umbrales de intensidad: 0 (no se quita ruido), 50, 100, 200, 500 y 1000 cuentas. La escala de color representa $\log_{10}(\text{intensidad})$ 182

Figura 50. Espectrogramas de baja resolución (1 s, 1 Da) para el rango de masas [200 Da –700 Da], aplicando el algoritmo de reducción de ruido con distintos umbrales (de izquierda a derecha y de arriba abajo), espectrograma original (con umbral 0), y espectrogramas con umbral 100, 500 y 1000. Corresponde a un extracto fenólico de pimiento. 183

Figura 51. *Clustering* rectangular aplicado a una muestra simple formada por 5 estándares comerciales de polifenoles. Arriba: cromatograma BPC. Abajo: representación del área de los *clusters* rectangulares, calculada con una ventana rectangular de anchura $t_R = 1$ s y altura $m/z=1$ Da. El área viene representada con una escala de color. La escala representa el logaritmo en base 10 del área..... 187

Figura 52. Detalle del *clustering* rectangular de la Figura 51 aplicado a una muestra simple. Se representa el área de los *clusters* rectangulares, con anchura de 1 s y altura de 1 Da. El área viene representada con una escala de color..... 187

Figura 53. Ejemplo de aplicación del método de *preclustering*. A la izquierda el espectrograma de baja resolución (1 s, 1 Da) de un extracto de pimiento con reducción de ruido de umbral 100. El cuadrado



rojo indica la zona a estudiar que comprende los intervalos en tiempo [20 min, 35 min] y masa [446 Da, 450 Da] y se representa a la derecha su espectrograma de alta resolución (1 s, 10 mDa)..... 191

Figura 54. *Preclustering*: fase 1, basado en cajas de $T_{\text{resol}}= 300$ s, $dt_{\text{r}}= 20$ s y $dm/z= 20$ mDa. Aplicado a la zona de estudio del espectrograma del extracto de pimienta de la Figura 53 (izquierda). 192

Figura 55. *Preclustering*: fase 2 basado en cajas de $T_{\text{resol}}=20$ s, $dt_{\text{r}}=20$ s, $dm/z=20$ mDa, sobre el *cluster* 1 de la Figura 54..... 193

Figura 56. Resultado de aplicar el *clustering* de puntos sobre los *clusters* 6 y 9 de la Figura 55, que habían quedado sin dividir por el *preclustering* fase 1 y 2. Se ha aplicado umbrales de $dt_{\text{r}}=10$ s y $dm/z=10$ mDa. 194

Figura 57. *Clustering* por detección de picos cromatográficos: resultado de aplicar la detección de picos cromatográficos a una parte del *cluster* 3 de la Figura 56, que no había podido dividirse por *preclustering* fases 1 y 2, o por *clustering* de puntos. 195

Figura 58. Resultado de aplicar el *preclustering* basado en cajas dos veces con $T_{\text{resol}}=300$ s y después $T_{\text{resol}}=20$ s, y el método de *clustering* de puntos. Se ha realizado un zoom centrado en $m/z=377$ Da del espectro de masas de un extracto fenólico de aceite de oliva (A). Método de *clustering* de detección de picos cromatográficos para el mismo caso que A (B). Espectrograma para la zona estudiada y separación de los distintos *clusters* por el método de detección de picos cromatográficos (C). Representación esquemática de los *clusters* finales (D). Representación esquemática de los *clusters* finales filtrados en área $>10^5$ (E). 198

Figura 59. Arriba: representación de todos los *clusters* obtenidos, para la muestra estudiada de pimienta, sin eliminar ninguno. En medio: representación de todos los *clusters* obtenidos filtrados en área $> 10^4$, y abajo: los *clusters* con área $> 10^5$ 203

Figura 60. A la izquierda: tabla con el número de *clusters* generados en cada una de las etapas del método de *clustering* de líneas. A la derecha: tabla con el tiempo computacional de cada etapa del *clustering* de líneas e histograma del área de los *clusters* generados. Corresponde a la zona de estudio comprendida entre los intervalos de tiempo [20 min, 35 min] y masa [446 Da, 450 Da] del extracto vegetal de pimienta, una vez que se ha quitado el ruido y se ha aplicado un umbral de intensidad de 100 cuentas. 205

Figura 61. Representación de 4 huellas espectrométricas: ión principal y sus variantes isotópicas, una en cada figura. 207

Figura 62. Ejemplo de una muestra simple mezcla de 5 estándares. A la izquierda: tabla con el número de *clusters* generados en cada una de las etapas del método de *clustering* de líneas. Se ha realizado con datos a los que se les ha eliminado el ruido y se ha puesto un umbral de intensidad de 100 cuentas . A la derecha: tabla con el tiempo computacional de cada etapa del *clustering* de líneas e histograma del área de los *clusters* generados. 209

Figura 63. Mismo ejemplo que la Figura 62 pero eliminando ruido generado por el proceso de *clustering*. 211



Figura 64. Representación de los *clusters* generados mediante el *clustering* de líneas, para una muestra compuesta de 5 estándares y filtrada para área > 10⁴.....212

Figura 65. Cromatograma BPC y espectrograma de la muestra simple compuesta por 5 estándares. Aparecen 8 huellas espectrométricas que se corresponden con los 5 compuestos de la mezcla más tres compuestos nuevos: 463.09 Da y t_r=6.9 min; 191.0569 Da y t_r=10.0 min; 331.04 Da y t_r=13.8 min.....217

Figura 66. Arriba: cromatograma BPC para la muestra 15. Centro: coeficiente de correlación correspondiente a las 20 muestras seleccionadas para entrenamiento $r(t_i, m_j)$ entre el área $A_n(t_i, m_j)$ y la actividad y_n en función del tiempo de retención y la m/z. Abajo: representación de la función $r_{12}(t_i)$: media de los 12 valores mayores de $r(t_i, m_j)$ en el tiempo de retención t_i232

Figura 67. Identificación de los candidatos a compuestos activos. Se genera un punto cuando $r_{12}(t_i) > 0.7$ y el más abundante de los 12 picos seleccionados también muestra un coeficiente de correlación mayor de 0.7. En las gráficas se representa m/z (arriba), el área (centro) y el coeficiente de correlación (abajo) de los picos espectrales más abundantes en función del tiempo de retención cuando se cumplen las condiciones impuestas.234

Figura 68. Representación de los coeficientes de correlación ($r > 0.5$) obtenidos utilizando la correlación entre el área de los compuestos obtenidos con *clustering* de líneas y la actividad de las 20 muestras de entrenamiento del experimento.237

Figura 69. Actividad (absorbancia) en función del área cromatográfica A_n para el primer candidato (m/z=595.1306 Da; t_r=6.25 min). Los círculos blancos representan las muestras de entrenamiento; los círculos negros corresponden a las muestras de evaluación. Las líneas sólida es la recta de regresión mientras que la línea punteada representa los límites del intervalo de confianza del 95% (estimado con las muestras de entrenamiento).239

Figura 70. Esquema de los estudios realizados en el bloque 1.248

Figura 71. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método TEAC.....261

Figura 72. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método FRAP.....262

Figura 73. Capacidad antioxidante de las frutas (A) y vegetales (B) mediante el método ORAC.....263

Figura 74. Cromatogramas de pico base de los tres extractos de pimienta analizados mediante HPLC-ESI-TOF-MS: A (pimiento verde), B (pimiento amarillo) y C (pimiento rojo).....274

Figura 75. Arriba: cromatogramas BPC y TIC de una muestra de pimiento verde. Abajo: Espectrograma de baja resolución (1 s, 1 Da) de la misma muestra.....288

Figura 76. Izquierda: detalle de una zona del espectrograma de la Figura 75 para el intervalo de tiempo [28.8 min, 30 min] y de masa m/z [429 Da, 465 Da]. Derecha: detalle de una zona del espectrograma de la Figura 75 para el intervalo de tiempo [29.68 min, 30 min] y de masa m/z [431 Da, 439 Da].288

Figura 77. Izquierda: número de *clusters* generados en cada fase del *clustering* de líneas. Derecha: histograma del área de los *clusters* y tiempo computacional invertido para la muestra de pimiento verde.....291



Figura 78. Representación de los *clusters* generados mediante la técnica de *clustering* lineal para la muestra de pimiento verde.292

Figura 79. Representación de los *clusters* generados mediante la técnica de *clustering* lineal para el detalle de la zona de la Figura 78.293

Figura 80. Cromatograma BPC del pimiento verde (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas (en rojo).295

Figura 81. Diferentes detalles de la Figura 80. BPC (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas (en rojo).296

Figura 82. Cromatograma BPC del pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado sólo los picos con intensidad $> 10^5$ cuentas (59 picos en rojo).296

Figura 83. Izquierda: detalle de la Figura 82 correspondiente a un pico con intensidad superior a 10^5 cuentas oculto bajo otro pico de mayor intensidad. Derecha: mismo detalle pero representando todos los picos encontrados por el *clustering* (detalle de la Figura 80).297

Figura 84. Cromatograma BPC del extracto de pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado sólo los picos con intensidad comprendida entre 10^5 y 5×10^4 cuentas (37 picos en rojo).297

Figura 85. Cromatograma BPC del pimiento verde (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 2 \times 10^5$ cuentas. Se han representado los picos con intensidad $< 5 \times 10^4$ cuentas (68 picos en rojo).298

Figura 86. Correlogramas correspondientes a los estudios de correlación entre los datos LC-MS y la capacidad antioxidante de los tres métodos TEAC, FRAP, y ORAC para las muestras de pimiento.304

Figura 87. Esquema de los estudios realizados en el bloque 2.313

Figura 88. Cromatograma de iones totales de la muestra de *Cistus ladanifer* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.315

Figura 89. Cromatograma de iones totales de la muestra de *Cistus albidus* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.315

Figura 90. Cromatograma de iones totales de la muestra de *Cistus salviifolius* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.316



Figura 91. Cromatograma de iones totales de la muestra de *Cistus clusii* analizada mediante SPME-GC-MS, representando algunos compuestos volátiles mayoritarios con su correspondiente intensidad y estructura.316

Figura 92. Gráficas de supervivencia de *E. coli* (A) y *S. aureus* (B) en presencia de concentraciones crecientes de los extractos de *C. salviifolius* S2 (extracto acuoso secado con spray-drying) y S2a (fracción fenólica derivada de S2).338

Figura 93. Cromatogramas de pico base de los extractos de *C. salviifolius*. (A) S2, (B) S4, (C). S2a.....340

Figura 94. Espectrograma de baja resolución (1 s, 1 Da) del extracto acuoso de *C. salviifolius* (S4) analizado mediante UPLC-QTOF-MS. Con el mapa de color se representa la intensidad para cada par ($t_R, m/z$).346

Figura 95. Zoom del espectrograma de baja resolución (1 s, 1 Da) realizado con el extracto S4. Con el mapa de color se representa la intensidad para cada par (t, m).....347

Figura 96. Clusters generados con el método de *clustering* de líneas a la muestra de *Cistus* S4. Se ha aplicado un umbral de área de 10^4 cuentas.350

Figura 97. Zoom de la representación de los *clusters* obtenidos para el extracto S4 tras aplicar la herramienta de *clustering* de picos con un umbral de área de 5×10^4 351

Figura 98. Cromatograma BPC del *Cistus* S4 (en azul) y todos los picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (en rojo).352

Figura 99. Diferentes detalles de la Figura 98. BPC (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (en rojo).353

Figura 100. Cromatograma BPC del *Cistus* S4 (en azul) y picos representantes de las familias de líneas encontrados mediante el *clustering* de líneas para un umbral de área $> 1 \times 10^4$ cuentas (picos en rojo). Se han representado tres gráficas correspondientes a los picos con intensidad $> 5 \times 10^4$ cuentas (35 picos), los picos con intensidad comprendida entre 5×10^4 y 1×10^4 cuentas (95 picos) y los picos con intensidad $< 1 \times 10^4$ cuentas (406 picos).....354

Figura 101. Cromatogramas de ion extraído (EIC) bajo el cromatograma de pico base (BPC) de los picos ocultos identificados gracias a la aplicación del *clustering* de picos. Corresponde con el extracto S4.356

Figura 102. Patrón de fragmentación para el compuesto tetrámero de galocatequina (GC4).360

Figura 103. Patrón de fragmentación para (A) coniferina, (B) quercetina glicósido galato y (C) elálgico-7-xilósido metilado.362

Figura 104. Extractos polifenólicos de *Cistus* utilizados para el estudio de correlación con la bioactividad.364

Figura 105. Cromatogramas BPC de los extractos de *Cistus*. A: 1 (rojo), 2 (verde), 3 (azul), 4 (amarillo), 5 (rosa). B: 6 (rojo), 7 (verde), 8 (azul), 9 (amarillo), 10 (rosa). C: 11 (rojo), 12 (verde), 13 (azul), 14



(amarillo), 15 (rosa). D: 16 (rojo), 17 (verde), 18 (azul), 19 (amarillo), 20 (rosa). E: 21 (rojo), 22 (verde), 23 (azul), 24 (amarillo), 25 (rosa) y 26 (gris).....365

Figura 106. Capacidad antimicrobiana de los 26 extractos de *Cistus* frente a bacterias Gram positivas y Gram negativas.....366

Figura 107. Correlograma realizado con los datos de MS de los 26 extractos de *Cistus* y los datos de capacidad antimicrobiana frente a *E. coli*. En escala de color se representa el coeficiente de correlación para cada *cluster* de los obtenidos con el *clustering* de picos. Resolución temporal: 1 s y resolución en masa: 1 Da.367

Figura 108. Correlograma realizado con los datos de MS de los 26 extractos de *Cistus* y los datos de capacidad antimicrobiana frente a *S. aureus*. En escala de color se representa el coeficiente de correlación para cada *cluster* de los obtenidos con el *clustering* de picos. Resolución temporal: 1 s y resolución en masa: 1 Da.368

Figure 109. Bioactivity of the isolated fractions from a polyphenolic olive oil extract (Picual/Arbequine) through *in vitro* studies against SKBR3 breast cancer cells.380

Figure 110. Outline studies in section 3.381

Figure 111. 3D spectrogram representation of one of the olive oil phenolic extract of the study.....382

Figure 112. BPC and TIC chromatograms and low resolution spectrogram (1 s, 1 Da) of one of the olive oil phenolic extracts of the study. For each interval of elution time and m/z ratio, the intensity is represented according to the colour scale.....382

Figure 113. Detail of the zone corresponding to m/z 377 Da in the spetrogram of the figure 112.383

Figure 114. MS data obtained by applying the line-*clustering* method with one of the olive oil phenolic extract of the study. Threshold area > 10⁵. A detail at right.384

Figure 115. Base-peak chromatogram (blue) of one olive oil sample and all the peaks depicting the families of lines (red) found through the line-*clustering*, area threshold of 1x10⁵.385

Figure 116. Base-peak chromatogram (blue) of the olive oil sample and all the peaks depicting the families of lines (red) found through the line-*clustering*, area threshold of 1x10⁵ and removal of peaks eluting continuously.385

Figure 117. Base-peak chromatogram (blue) of the olive oil sample and all the peaks depicting the families of lines (red) found through the line-*clustering*, area threshold of 1x10⁵. (A) intensity peaks > 10⁵ (54 peaks), (B) 10⁵ > Intensity > 10⁴ (257 peaks), and intensity peaks < 10⁴ (122 peaks).386

Figure 118. Description of sample 2 in experiment involving EVOO-PE. Top plot: base peak chromatogram. Bottom plot: representation of the area as a function of retention time and m/z. HPLC-MS data are represented after bucketing using $A_2(t_i, m_j)$ with resolutions of 1 s and 1 Da. Area represented in logarithmic scale according to the colour scale.....390

Figure 119. Correlation coefficient for the collection of samples in the training partition. The top plot contains the base peak chromatogram (for sample 2). In the middle, the correlation coefficient $r(t_i, m_j)$ between the area $A_n(t_i, m_j)$ and the bioactivity y_n is represented as a function of the elution time and



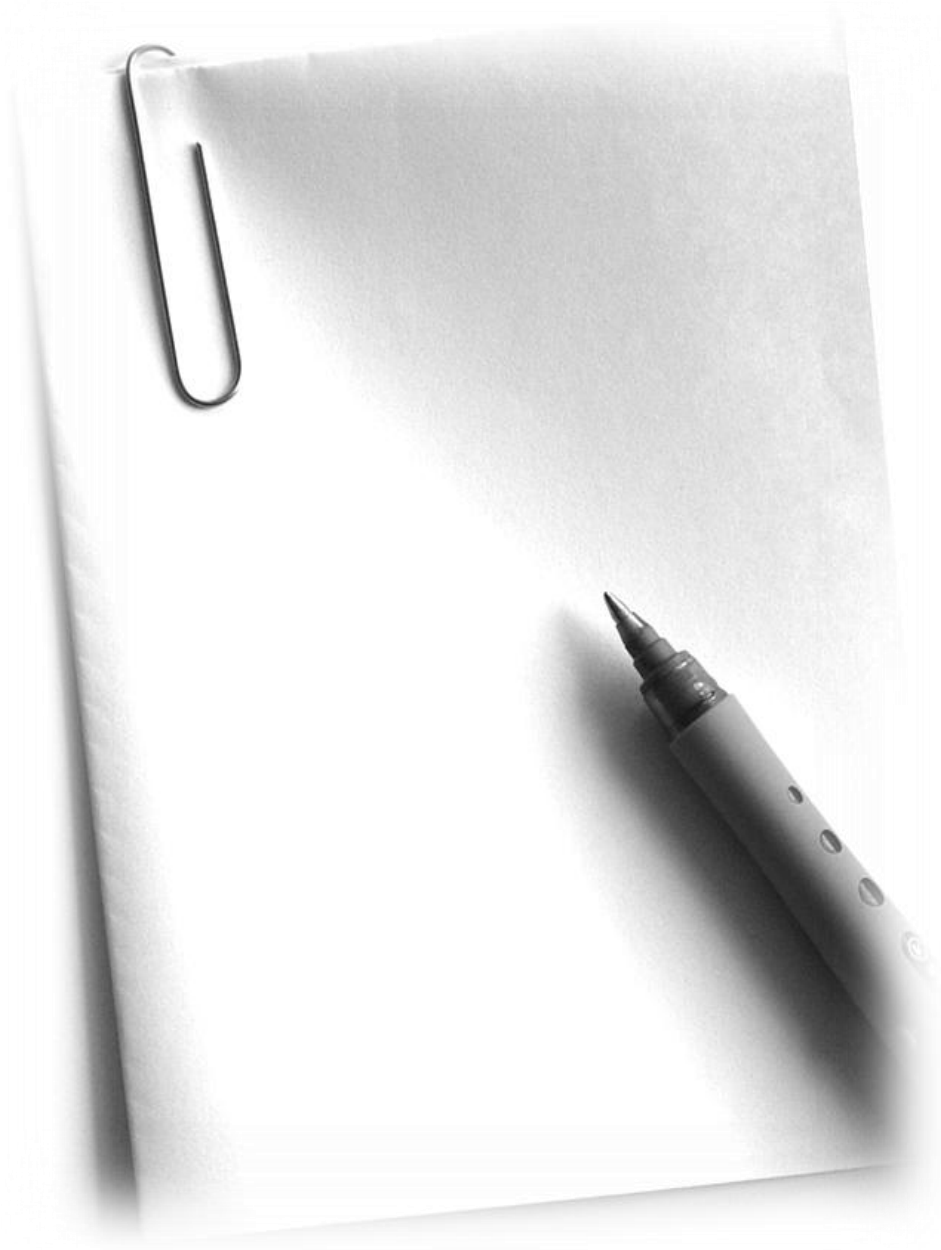
the m/z ratio. The plot at the bottom represents the function $r_{12}(t_i)$ average of the best 12 correlation values of $r(t_i, m_j)$ at elution time t_i391

Figure 120. Identification of candidates for the bioactive compound. A point was generated in the plots when $r_{12}(t_i) > 0.85$ and the most abundant of the 12 selected spectral peaks also showed a correlation coefficient greater than 0.85. The plots represent the m/z ratio (top), the area (middle) and the correlation coefficient (bottom) of this most abundant spectral peak, as a function of elution time when the above conditions are satisfied.393

Figure 122. Correlation coefficient for the collection of 22 olive oil polyphenolic extracts. The correlation coefficient $r(t_i, m_j)$ between the area $A_n(t_i, m_j)$ and the bioactivity y_n is represented as a function of the retention time and the m/z ratio.394

Figure 121. Bioactivity (cytotoxicity against JIMT-1 breast cancer) as a function of the chromatographic area A_n for candidate 6 (m/z ratio: 377.1252 Da; elution time: 1254 s). Unfilled and filled circles represent the samples in the training and evaluation partitions, respectively. The solid and dotted lines are the regression line and the limits of the 95% confidence interval (estimated from the training partition).....397





APÉNDICE

Este apéndice se divide en tres partes, correspondientes a las muestras estudiadas en cada bloque de la presente tesis. Es decir, el apéndice 1 está dedicado a las muestras de pimiento, el apéndice 2 a las muestras de *Cistus* y el apéndice 3 a las muestras de aceite.

En cada apéndice, se indican en primer lugar los parámetros utilizados en el *clustering* de líneas. Después, se muestran los resultados de la aplicación del *clustering* de líneas para una muestra representante de cada grupo (pimiento, *Cistus* y aceite). Para la muestra representante de cada bloque, se indica en las tablas el número de *clusters* generado y el tiempo computacional de cada fase del *clustering* de líneas y se presenta una gráfica con el histograma del área de los *clusters* generados, para verificar que el umbral aplicado en la eliminación de ruido es el adecuado.

Para la muestra representante de cada grupo, se presenta el listado de compuestos resultante del *clustering* de líneas, estableciendo un umbral de área determinado y ordenado por tiempo de retención. Se indica: el índice del compuesto (asignado por el programa), el tiempo de retención (t_R en minutos) y desviación en tiempo, la masa exacta experimental (m/z en Da) y desviación en masa (en mDa), el área, y el número de líneas que componen la huella espectrométrica del compuesto. Para cada compuesto, se presentan las 10 líneas de mayor área con sus porcentajes relativos al área de la línea principal.

En los resultados del *clustering* de líneas también se genera de manera automática una lista de comparación de las áreas de todos los *clusters* obtenidos en todas las muestras (3 muestras en el estudio del pimiento, 26 muestras en el estudio del *Cistus* y 22 muestras en el estudio del aceite). Se presenta esta tabla de comparación entre muestras únicamente para el caso del pimiento ya que las listas generadas para las colecciones de muestras de *Cistus* y aceite son demasiado extensas.



APÉNDICE 1: PIMIENTO.

Se ha elegido una muestra representante del grupo de muestras de pimiento con las siguientes características:

Nombre de la muestra	pimiento_v1
Intervalo masas analizadas	[100 Da, 900 Da]
Intervalo de tiempo analizado	[0 min, 30 min]
Tamaño fichero	26MB

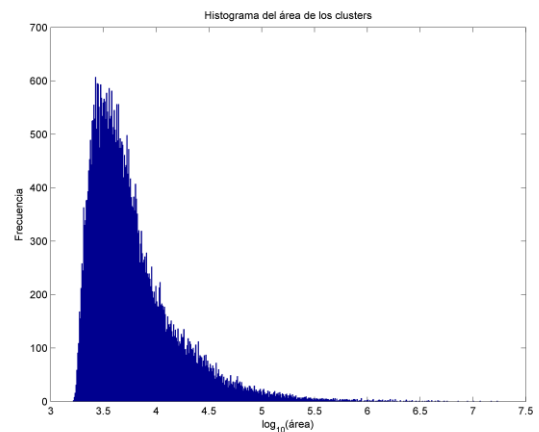
Se ha realizado la transformación del formato baf a XML, después se ha convertido en formato binario, se ha eliminado el ruido y se ha calibrado en masa. Se ha realizado el *clustering* de líneas, utilizando los siguientes parámetros:

Nombre del parámetro	Valor utilizado
Umbral intensidad	400 cuentas
Elimina <i>clusters</i> : Umbral área	1000
Elimina <i>clusters</i> : N_puntos	3
Resolución en tiempo fase 1	300s
Resolución en tiempo fase 2	20s
Umbral tiempo de retención	10s
Umbral m/z	0.010Da
Umbral área huella	10 ⁴



Tablas con el número de *clusters* y el tiempo computacional, y el histograma del área de los *clusters* obtenidos con el clustering de líneas:

<i>Clustering</i> de líneas	Nº de <i>clusters</i>
Número inicial de puntos (t,m/z,l) sin ruido	1457204
Número de puntos (t,m/z,l) sin ruido, umbral_l =400	404923
<i>Pre-Clustering</i> fase1: resolución tiempo =300s	11607
<i>Pre-Clustering</i> fase2: : resolución tiempo =20s	40507
Elimina: puntos<3, area<1000	10932
<i>Clustering</i> de puntos	114494
Elimina: puntos<3, area<1000	30188
Detección de picos cromatográficos	61291
Elimina: puntos<3, area<1000	48227
Umbral de area> 10 ⁴	10662
Umbral de area> 10 ⁵	948
Umbral de area> 10 ⁶	92



Tiempo computacional

<i>Preclustering</i> : fase 1 y 2	8 min	24%
<i>Clustering</i> de puntos	11.4 min	34.6%
Detección de picos cromatográficos	9.3 min	28.3%
Estadística	4.3 min	13.1%
TOTAL	33 min	100%



Parte del listado de compuestos resultados del *clustering* de líneas. El parámetro *Cluster* es el índice del *cluster*, el área se expresa en Kcuentas, la masa y su desviación estándar en Da y mDa respectivamente, la máxima diferencia entre las masas de los puntos que componen los *clusters* (rangoM) en mDa, el tiempo de retención y su desviación estándar en segundos y la diferencia máxima de los t_R de los puntos que componen el *cluster* (rangoT) en segundos.

%Cluster	Area	mass	SE_M	rangoM	time	SE_T	rangoT
%	(kCuentas)	(Da)	(mDa)	(mDa)	(s)	(s)	(s)
1204	17124.0000	191.0589	0.5364	3.5400	126.6911	6.5605	29.110
38129	14848.1360	447.1006	2.6429	19.5007	1597.8287	5.2489	47.170
43015	11752.3120	579.1410	2.0681	23.6206	1435.2360	6.4235	59.210
16692	10180.1280	581.1531	1.3096	13.6108	1080.6409	11.7912	68.240
44611	8808.9080	621.1508	1.7124	20.9351	1613.9444	5.7244	55.190
45827	7298.2520	665.1408	1.4897	19.5923	1613.5009	5.1839	46.160
44151	5641.0840	611.1464	0.9702	6.1646	359.9403	23.6593	101.360
48144	5418.4320	187.0994	0.5718	9.0637	1647.4065	3.0642	36.130
17349	5328.6440	593.1561	1.8496	15.2588	1146.5116	3.3820	32.120
38147	4818.1600	448.0987	0.6722	7.3547	1597.6122	4.6798	35.130
6295	4728.9880	191.0212	0.8346	21.0114	201.6762	15.7109	65.230
44185	4646.1520	611.1647	1.4822	18.0664	989.0769	8.4994	67.240
29772	4637.8680	341.0915	0.6157	7.6599	1279.1950	15.0377	71.250
1100	4390.2480	895.2018	1.9046	14.4653	1596.9954	3.4120	18.060
43069	4073.4360	580.1396	0.7158	9.2773	1434.7007	5.5960	44.160
15861	4072.8000	563.1448	1.7869	21.4233	1223.1392	2.9439	20.070
14680	3875.4440	539.1791	1.0854	27.8931	1733.5345	3.3383	44.160
42430	3807.8000	563.1449	1.2259	20.8740	1321.4673	3.0539	31.110
42427	3772.3520	563.1420	0.5227	10.0098	1258.1897	4.5496	32.110
26379	3712.7200	282.0836	0.3904	10.3455	505.8152	7.9864	62.220
47430	3633.6120	363.0730	0.6167	21.8201	1356.5166	7.4152	51.180
16744	3510.8800	582.1551	0.8222	6.5918	1080.1552	11.6914	65.220
38118	3357.3160	447.0958	0.8158	28.5034	1310.3494	4.4021	62.220
44629	3085.6960	622.1517	0.8932	18.6768	1613.6650	5.2590	42.150
17526	2945.5000	597.1484	1.2163	22.6440	968.5656	10.8775	86.310
38117	2894.6960	447.0948	0.8468	11.6272	1270.9004	3.2618	32.110
2724	2679.5080	651.1969	0.9653	18.7988	977.9182	2.2731	20.070
45865	2669.4960	666.1417	0.8175	24.9634	1613.4363	5.1431	44.160



1233	2625.1080	203.0843	0.4313	9.8114	769.1273	5.9315	53.190
15931	2528.1440	565.1760	0.6314	10.1929	507.9011	5.9504	36.130
43004	2460.4240	579.1378	1.1004	17.6392	1208.0977	3.3871	36.130
5682	2399.8520	164.0730	0.3632	11.5204	550.6988	5.0020	38.140
47940	2357.8000	243.1243	0.5393	13.6871	1673.0920	3.5056	39.130
32033	2341.0840	191.0213	0.3595	1.6785	247.4154	6.6966	23.080
43000	2325.5680	579.1375	1.0524	30.8228	1139.2161	2.7564	29.100
26544	2283.5080	290.0881	0.5315	9.1553	304.9317	12.4103	68.240
1101	2218.5320	896.2024	1.0616	10.5591	1596.9239	3.3857	18.060
44091	2058.4040	609.1483	0.9066	13.4277	1060.0056	2.9118	26.090
10533	2053.6800	433.1149	0.5346	10.7422	1795.3030	2.3329	13.040
47652	1908.6320	761.3250	1.0234	5.1270	1745.4703	6.1521	26.100
17545	1842.0440	597.1841	0.7015	14.4653	1446.0335	2.8589	27.100
7816	1835.2760	325.0938	0.3737	8.6670	978.0862	2.9749	35.130
29755	1826.2000	341.0888	1.5724	21.2708	971.0190	3.9112	38.140
43597	1824.2400	594.1563	1.0421	26.4893	1146.3873	3.3679	40.140
2159	1806.7160	461.1091	0.6940	16.5710	1767.0869	3.1592	31.110
47113	1754.3760	741.1893	0.8784	15.6250	1106.1829	2.9926	38.130
25	1741.3760	145.0628	0.4065	6.0120	126.2524	7.3015	48.170
44153	1724.8040	611.1452	0.8490	13.7329	414.9076	4.2768	32.120
27114	1651.5400	191.0568	0.5568	4.4098	148.7347	3.7990	15.050
26780	1649.3760	180.0679	0.3454	2.4719	370.5434	7.5053	35.120
11562	1631.9600	463.0898	0.9621	18.5242	1462.0388	3.2234	26.090
44610	1631.6920	621.1475	0.9459	18.1885	1580.6046	4.6715	29.110
44093	1594.4800	609.1470	0.9508	24.4751	1102.2983	3.6466	29.100
44233	1591.8640	612.1671	1.9378	28.0762	988.9567	8.6692	67.240
28267	1570.4840	612.1489	1.7130	19.4092	357.7409	26.3400	125.440
53	1546.3200	192.0603	0.3173	2.3956	126.7229	6.4469	29.100
44189	1524.4640	611.1624	0.9848	11.8408	1124.4799	8.0703	44.150
180	1513.8200	316.9470	0.3634	2.1057	22.6036	3.1604	13.050
11028	1481.0800	447.0941	0.6621	16.2048	1464.5719	3.7612	32.110
26543	1474.0560	290.0880	0.3561	1.2512	280.7222	5.5302	20.070
15973	1453.4920	565.1575	0.8495	14.7705	1571.3009	2.9518	26.090
31615	1403.9320	199.0261	0.4975	7.1411	1355.4814	7.7177	56.200
41693	1392.9800	545.2606	0.9650	19.5313	1776.4034	3.6907	30.110



32032	1392.4960	191.0212	0.1922	0.5341	230.3494	3.4541	11.040
15913	1356.2000	564.1454	0.5856	5.3101	1222.9288	2.5766	16.060
10919	1342.7760	445.1957	0.5463	11.3525	792.4166	3.4622	26.090
32034	1333.6560	191.0219	1.9463	19.5465	270.9827	6.8959	41.140
29159	1322.3360	243.0623	0.6810	26.5808	395.2926	3.8812	34.120
28570	1321.3320	163.0416	0.3581	10.2386	978.5502	3.8355	40.150
10442	1297.8400	431.0986	0.6167	10.1624	1750.1348	2.9660	23.080
35357	1295.4360	359.1487	1.7933	24.9023	1364.7651	3.3887	36.130
257	1292.7560	384.9351	0.5036	4.8218	22.7038	3.1346	13.050
42461	1276.7200	564.1461	1.0561	26.7944	1321.4629	3.0436	27.090
29771	1266.2720	341.0917	0.5551	9.6436	1239.4368	6.5596	32.120
396	1241.0360	520.9100	0.6505	2.5024	22.8247	3.0072	12.050
47653	1224.8120	761.3238	1.4353	10.6812	1771.9331	8.0157	40.140
34547	1197.7960	326.1245	1.2254	27.2522	591.6486	13.6311	65.220
26542	1166.7440	290.0880	0.5304	3.4790	257.0071	7.5519	29.100
44041	1147.0440	607.1676	1.2120	26.0620	1743.3380	4.1093	31.110
99	1142.5360	248.9603	0.5023	4.3945	22.5711	3.2957	15.060
323	1141.8280	452.9228	0.5734	2.2888	22.7441	3.0457	13.050
456	1127.4840	588.8973	0.5945	2.3804	22.9314	2.9264	11.040
47428	1118.0640	363.0761	0.7927	8.3923	1273.4984	9.7574	47.160
40754	1116.0600	521.2032	1.0504	18.0054	1493.9148	3.8418	35.130
42459	1100.6920	564.1441	0.9455	14.4043	1258.1813	4.4694	27.090
16702	1079.2600	581.1410	1.3474	25.5737	1434.3503	5.0789	29.100
41496	1051.8720	540.1806	1.0134	21.3623	1733.4969	3.2771	28.100
28625	1047.9760	155.0366	0.4022	11.1542	1355.4468	7.7642	47.170
26673	1024.7720	439.0849	0.9211	8.3618	142.5912	7.3321	40.140
44755	1017.6480	627.1580	1.7028	36.1328	889.4193	7.1612	56.200
15868	1011.9280	563.1420	0.7097	20.3247	1570.5807	3.3793	36.120
26538	1007.9040	290.0870	0.6314	3.0823	185.5433	9.6222	35.130
33674	998.2640	292.1391	0.7267	13.0920	421.6271	9.5530	50.180
11088	979.1640	449.1001	1.2770	22.1558	1597.5470	4.9350	40.150
44152	977.2640	611.1443	0.4287	1.4648	401.1864	3.2858	11.040
28990	951.8920	201.0262	0.3397	8.3771	102.5769	2.3832	24.080
20284	947.8800	652.1990	0.6822	4.3335	977.8717	2.2013	13.050
31431	924.7120	164.0726	0.5710	5.8289	591.8429	14.1134	65.230



Lista de compuestos obtenidos con el *clustering* de líneas para la muestra representativa de pimienta:

Muestra: 1. Fichero con clusters: cluster_estad_pimiento_v1_eliminal000.m
 Umbral area: 200000 cuentas
 Eliminados 157 clusters alrededor de la masa 283.2641 Da: t_tot=493.6 s de 1794.6 s
 Eliminados 146 clusters alrededor de la masa 255.2328 Da: t_tot=472.5 s de 1794.6 s

Numero de clusters iniciales: 11039

Lista de compuestos ordenados por tiempo de retención

Comp.1	m= 316.9470 Da	se(m)= 0.4 mDa	t= 0.38 min	t= 22.6 s	se(t)= 3.2 s	a= 1513.8 Kcuentas	tiene 120 líneas
Comp.2	m= 201.0262 Da	se(m)= 0.3 mDa	t= 1.71 min	t=102.6 s	se(t)= 2.3 s	a= 948.3 Kcuentas	tiene 119 líneas
Comp.3	m= 194.9452 Da	se(m)= 2.5 mDa	t= 1.83 min	t=109.8 s	se(t)= 1.1 s	a= 254.1 Kcuentas	tiene 85 líneas
Comp.4	m= 191.0589 Da	se(m)= 0.5 mDa	t= 2.11 min	t=126.7 s	se(t)= 6.6 s	a= 17124.0 Kcuentas	tiene 65 líneas
Comp.5	m= 307.1136 Da	se(m)= 0.6 mDa	t= 2.20 min	t=132.0 s	se(t)= 4.7 s	a= 661.0 Kcuentas	tiene 31 líneas
Comp.6	m= 341.1087 Da	se(m)= 0.6 mDa	t= 2.32 min	t=138.9 s	se(t)= 2.6 s	a= 811.0 Kcuentas	tiene 65 líneas
Comp.7	m= 439.0850 Da	se(m)= 0.9 mDa	t= 2.37 min	t=142.3 s	se(t)= 7.0 s	a= 1004.3 Kcuentas	tiene 61 líneas
Comp.8	m= 191.0568 Da	se(m)= 0.6 mDa	t= 2.48 min	t=148.7 s	se(t)= 3.8 s	a= 1651.5 Kcuentas	tiene 109 líneas
Comp.9	m= 133.0148 Da	se(m)= 0.5 mDa	t= 2.68 min	t=160.7 s	se(t)= 7.0 s	a= 524.9 Kcuentas	tiene 41 líneas
Comp.10	m= 341.1082 Da	se(m)= 0.8 mDa	t= 2.83 min	t=170.0 s	se(t)= 8.1 s	a= 433.4 Kcuentas	tiene 13 líneas
Comp.11	m= 290.0870 Da	se(m)= 0.6 mDa	t= 3.10 min	t=186.3 s	se(t)= 9.6 s	a= 1016.5 Kcuentas	tiene 26 líneas
Comp.12	m= 191.0212 Da	se(m)= 0.8 mDa	t= 3.36 min	t=201.7 s	se(t)=15.7 s	a= 4729.0 Kcuentas	tiene 1 líneas
Comp.13	m= 290.0876 Da	se(m)= 0.4 mDa	t= 3.49 min	t=209.2 s	se(t)= 3.4 s	a= 305.4 Kcuentas	tiene 29 líneas
Comp.14	m= 503.1622 Da	se(m)= 1.1 mDa	t= 3.58 min	t=214.9 s	se(t)= 4.2 s	a= 638.3 Kcuentas	tiene 38 líneas
Comp.15	m= 217.0362 Da	se(m)= 0.6 mDa	t= 3.71 min	t=222.5 s	se(t)=12.6 s	a= 727.4 Kcuentas	tiene 2 líneas
Comp.16	m= 191.0212 Da	se(m)= 0.2 mDa	t= 3.84 min	t=230.3 s	se(t)= 3.5 s	a= 1392.5 Kcuentas	tiene 41 líneas
Comp.17	m= 405.0293 Da	se(m)= 0.9 mDa	t= 4.01 min	t=240.5 s	se(t)=11.5 s	a= 645.8 Kcuentas	tiene 5 líneas
Comp.18	m= 191.0213 Da	se(m)= 0.4 mDa	t= 4.12 min	t=247.4 s	se(t)= 6.7 s	a= 2341.1 Kcuentas	tiene 27 líneas
Comp.19	m= 290.0880 Da	se(m)= 0.5 mDa	t= 4.29 min	t=257.6 s	se(t)= 7.4 s	a= 1183.0 Kcuentas	tiene 32 líneas
Comp.20	m= 191.0218 Da	se(m)= 1.9 mDa	t= 4.52 min	t=271.0 s	se(t)= 6.9 s	a= 1332.7 Kcuentas	tiene 20 líneas
Comp.21	m= 290.0880 Da	se(m)= 0.4 mDa	t= 4.67 min	t=280.0 s	se(t)= 4.8 s	a= 1272.0 Kcuentas	tiene 42 líneas
Comp.22	m= 290.0881 Da	se(m)= 0.4 mDa	t= 5.05 min	t=303.1 s	se(t)=11.3 s	a= 2414.6 Kcuentas	tiene 20 líneas
Comp.23	m= 217.0369 Da	se(m)= 0.9 mDa	t= 5.29 min	t=317.5 s	se(t)= 5.9 s	a= 334.8 Kcuentas	tiene 31 líneas
Comp.24	m= 611.1462 Da	se(m)= 1.1 mDa	t= 5.45 min	t=326.9 s	se(t)=14.0 s	a= 1487.5 Kcuentas	tiene 9 líneas
Comp.25	m= 323.0293 Da	se(m)= 1.4 mDa	t= 5.73 min	t=343.8 s	se(t)= 5.8 s	a= 379.4 Kcuentas	tiene 37 líneas
Comp.26	m= 330.0482 Da	se(m)= 1.8 mDa	t= 5.97 min	t=358.1 s	se(t)= 9.3 s	a= 747.6 Kcuentas	tiene 22 líneas
Comp.27	m= 633.1279 Da	se(m)= 1.8 mDa	t= 6.09 min	t=365.3 s	se(t)=16.3 s	a= 618.8 Kcuentas	tiene 8 líneas
Comp.28	m= 611.1465 Da	se(m)= 0.8 mDa	t= 6.19 min	t=371.2 s	se(t)=12.8 s	a= 4082.7 Kcuentas	tiene 31 líneas



Comp.29	m= 362.0518	Da	se (m)= 1.6 mDa	t= 6.33 min	t=379.8 s	se (t)= 8.3 s	a= 475.3	Kcuentas	tiene 36 líneas
Comp.30	m= 174.0776	Da	se (m)= 1.0 mDa	t= 6.50 min	t=389.7 s	se (t)= 4.0 s	a= 266.2	Kcuentas	tiene 29 líneas
Comp.31	m= 243.0622	Da	se (m)= 0.5 mDa	t= 6.58 min	t=395.0 s	se (t)= 3.9 s	a= 1344.5	Kcuentas	tiene 108 líneas
Comp.32	m= 184.0622	Da	se (m)= 0.9 mDa	t= 6.59 min	t=395.1 s	se (t)= 9.7 s	a= 435.0	Kcuentas	tiene 2 líneas
Comp.33	m= 611.1443	Da	se (m)= 0.6 mDa	t= 6.69 min	t=401.4 s	se (t)= 4.1 s	a= 1167.1	Kcuentas	tiene 49 líneas
Comp.34	m= 773.1976	Da	se (m)= 1.4 mDa	t= 6.81 min	t=408.4 s	se (t)= 7.3 s	a= 384.3	Kcuentas	tiene 23 líneas
Comp.35	m= 611.1452	Da	se (m)= 1.0 mDa	t= 6.93 min	t=415.8 s	se (t)= 5.5 s	a= 1638.3	Kcuentas	tiene 73 líneas
Comp.36	m= 292.1391	Da	se (m)= 0.6 mDa	t= 7.02 min	t=421.3 s	se (t)= 9.6 s	a= 1018.1	Kcuentas	tiene 24 líneas
Comp.37	m= 773.1980	Da	se (m)= 1.9 mDa	t= 7.35 min	t=440.9 s	se (t)= 3.8 s	a= 227.9	Kcuentas	tiene 66 líneas
Comp.38	m= 360.0716	Da	se (m)= 1.6 mDa	t= 8.26 min	t=495.5 s	se (t)= 5.5 s	a= 290.0	Kcuentas	tiene 67 líneas
Comp.39	m= 282.0836	Da	se (m)= 0.4 mDa	t= 8.43 min	t=505.8 s	se (t)= 8.0 s	a= 3710.3	Kcuentas	tiene 65 líneas
Comp.40	m= 549.1811	Da	se (m)= 1.4 mDa	t= 8.47 min	t=508.4 s	se (t)= 2.3 s	a= 328.9	Kcuentas	tiene 65 líneas
Comp.41	m= 356.1272	Da	se (m)= 1.4 mDa	t= 8.69 min	t=521.3 s	se (t)= 5.1 s	a= 201.5	Kcuentas	tiene 44 líneas
Comp.42	m= 164.0730	Da	se (m)= 0.3 mDa	t= 9.18 min	t=550.8 s	se (t)= 5.1 s	a= 2410.0	Kcuentas	tiene 99 líneas
Comp.43	m= 536.1149	Da	se (m)= 2.3 mDa	t= 9.32 min	t=559.2 s	se (t)= 7.0 s	a= 327.3	Kcuentas	tiene 43 líneas
Comp.44	m= 299.0769	Da	se (m)= 0.8 mDa	t= 9.41 min	t=564.4 s	se (t)= 6.0 s	a= 243.5	Kcuentas	tiene 34 líneas
Comp.45	m= 236.0935	Da	se (m)= 0.8 mDa	t= 9.77 min	t=586.2 s	se (t)= 7.6 s	a= 316.5	Kcuentas	tiene 40 líneas
Comp.46	m= 326.1246	Da	se (m)= 1.0 mDa	t= 9.86 min	t=591.4 s	se (t)=13.2 s	a= 1189.6	Kcuentas	tiene 14 líneas
Comp.47	m= 231.0514	Da	se (m)= 0.9 mDa	t=10.00 min	t=600.2 s	se (t)= 7.8 s	a= 485.6	Kcuentas	tiene 34 líneas
Comp.48	m= 199.0262	Da	se (m)= 0.8 mDa	t=10.24 min	t=614.6 s	se (t)= 3.9 s	a= 274.4	Kcuentas	tiene 49 líneas
Comp.49	m= 259.1296	Da	se (m)= 1.0 mDa	t=10.39 min	t=623.5 s	se (t)= 7.0 s	a= 244.5	Kcuentas	tiene 37 líneas
Comp.50	m= 218.1041	Da	se (m)= 0.4 mDa	t=10.91 min	t=654.5 s	se (t)= 2.8 s	a= 274.3	Kcuentas	tiene 70 líneas
Comp.51	m= 288.0718	Da	se (m)= 0.7 mDa	t=11.00 min	t=660.2 s	se (t)= 4.1 s	a= 239.2	Kcuentas	tiene 34 líneas
Comp.52	m= 329.0873	Da	se (m)= 0.6 mDa	t=11.35 min	t=680.9 s	se (t)= 4.1 s	a= 457.5	Kcuentas	tiene 64 líneas
Comp.53	m= 586.1312	Da	se (m)= 2.4 mDa	t=11.96 min	t=717.8 s	se (t)= 3.5 s	a= 208.0	Kcuentas	tiene 41 líneas
Comp.54	m= 299.0773	Da	se (m)= 0.9 mDa	t=12.12 min	t=727.2 s	se (t)= 2.9 s	a= 220.1	Kcuentas	tiene 54 líneas
Comp.55	m= 442.1296	Da	se (m)= 1.2 mDa	t=12.31 min	t=738.8 s	se (t)= 5.3 s	a= 515.0	Kcuentas	tiene 47 líneas
Comp.56	m= 727.1951	Da	se (m)= 2.4 mDa	t=12.58 min	t=755.0 s	se (t)=10.3 s	a= 580.0	Kcuentas	tiene 20 líneas
Comp.57	m= 203.0843	Da	se (m)= 0.4 mDa	t=12.82 min	t=769.1 s	se (t)= 5.8 s	a= 2621.1	Kcuentas	tiene 85 líneas
Comp.58	m= 445.1957	Da	se (m)= 0.5 mDa	t=13.21 min	t=792.4 s	se (t)= 3.4 s	a= 1340.8	Kcuentas	tiene 50 líneas
Comp.59	m= 233.1299	Da	se (m)= 1.0 mDa	t=13.31 min	t=798.4 s	se (t)= 6.8 s	a= 202.7	Kcuentas	tiene 8 líneas
Comp.60	m= 403.1254	Da	se (m)= 0.5 mDa	t=13.39 min	t=803.1 s	se (t)= 2.5 s	a= 644.4	Kcuentas	tiene 55 líneas
Comp.61	m= 343.1034	Da	se (m)= 0.6 mDa	t=13.48 min	t=809.1 s	se (t)= 2.9 s	a= 453.1	Kcuentas	tiene 22 líneas
Comp.62	m= 759.2025	Da	se (m)= 4.0 mDa	t=13.83 min	t=829.6 s	se (t)= 7.9 s	a= 217.8	Kcuentas	tiene 24 líneas
Comp.63	m= 451.1818	Da	se (m)= 2.2 mDa	t=13.97 min	t=838.0 s	se (t)=13.2 s	a= 251.1	Kcuentas	tiene 2 líneas
Comp.64	m= 582.1908	Da	se (m)= 1.5 mDa	t=13.98 min	t=838.7 s	se (t)= 2.8 s	a= 389.1	Kcuentas	tiene 50 líneas
Comp.65	m= 597.1830	Da	se (m)= 2.4 mDa	t=14.08 min	t=844.7 s	se (t)= 6.5 s	a= 406.8	Kcuentas	tiene 36 líneas
Comp.66	m= 309.0611	Da	se (m)= 1.1 mDa	t=14.39 min	t=863.2 s	se (t)= 5.1 s	a= 603.9	Kcuentas	tiene 47 líneas
Comp.67	m= 263.1393	Da	se (m)= 1.3 mDa	t=14.58 min	t=875.1 s	se (t)= 7.7 s	a= 386.1	Kcuentas	tiene 34 líneas



Comp.68	m= 443.1930	Da	se(m)= 1.5 mDa	t=14.74 min	t=884.4 s	se(t)= 2.7 s	a= 387.9	Kcuentas	tiene 36 líneas
Comp.69	m= 627.1579	Da	se(m)= 1.3 mDa	t=14.82 min	t=889.4 s	se(t)= 6.9 s	a= 1011.0	Kcuentas	tiene 24 líneas
Comp.70	m= 320.1069	Da	se(m)= 0.9 mDa	t=15.18 min	t=910.6 s	se(t)= 6.3 s	a= 371.8	Kcuentas	tiene 33 líneas
Comp.71	m= 215.1409	Da	se(m)= 1.0 mDa	t=15.73 min	t=943.9 s	se(t)= 3.3 s	a= 475.4	Kcuentas	tiene 48 líneas
Comp.72	m= 597.1484	Da	se(m)= 1.2 mDa	t=16.14 min	t=968.7 s	se(t)=10.9 s	a= 2941.5	Kcuentas	tiene 12 líneas
Comp.73	m= 341.0887	Da	se(m)= 0.8 mDa	t=16.18 min	t=970.9 s	se(t)= 3.4 s	a= 1811.3	Kcuentas	tiene 54 líneas
Comp.74	m= 651.1969	Da	se(m)= 0.9 mDa	t=16.30 min	t=977.9 s	se(t)= 2.3 s	a= 2677.7	Kcuentas	tiene 77 líneas
Comp.75	m= 611.1647	Da	se(m)= 1.5 mDa	t=16.48 min	t=989.1 s	se(t)= 8.5 s	a= 4645.2	Kcuentas	tiene 23 líneas
Comp.76	m= 406.1667	Da	se(m)= 0.7 mDa	t=17.27 min	t=1036.4 s	se(t)= 2.7 s	a= 854.7	Kcuentas	tiene 73 líneas
Comp.77	m= 595.1676	Da	se(m)= 1.5 mDa	t=17.51 min	t=1050.6 s	se(t)= 3.7 s	a= 267.8	Kcuentas	tiene 76 líneas
Comp.78	m= 609.1483	Da	se(m)= 0.9 mDa	t=17.67 min	t=1060.0 s	se(t)= 2.9 s	a= 2058.4	Kcuentas	tiene 61 líneas
Comp.79	m= 465.1041	Da	se(m)= 1.1 mDa	t=17.76 min	t=1065.4 s	se(t)= 8.2 s	a= 671.8	Kcuentas	tiene 13 líneas
Comp.80	m= 581.1531	Da	se(m)= 1.3 mDa	t=18.01 min	t=1080.6 s	se(t)=11.8 s	a= 10178.7	Kcuentas	tiene 18 líneas
Comp.81	m= 611.1627	Da	se(m)= 1.7 mDa	t=18.10 min	t=1085.9 s	se(t)= 4.6 s	a= 256.8	Kcuentas	tiene 39 líneas
Comp.82	m= 741.1893	Da	se(m)= 0.9 mDa	t=18.44 min	t=1106.2 s	se(t)=2.9 s	a= 1751.4	Kcuentas	tiene 82 líneas
Comp.83	m= 673.3050	Da	se(m)= 0.8 mDa	t=18.54 min	t=1112.6 s	se(t)= 2.9 s	a= 823.8	Kcuentas	tiene 40 líneas
Comp.84	m= 611.1624	Da	se(m)= 1.0 mDa	t=18.74 min	t=1124.5 s	se(t)= 8.1 s	a= 1524.5	Kcuentas	tiene 19 líneas
Comp.85	m= 579.1375	Da	se(m)= 0.7 mDa	t=18.99 min	t=1139.2 s	se(t)= 2.7 s	a= 2317.7	Kcuentas	tiene 62 líneas
Comp.86	m= 593.1561	Da	se(m)= 1.8 mDa	t=19.11 min	t=1146.5 s	se(t)= 3.4 s	a= 5328.6	Kcuentas	tiene 33 líneas
Comp.87	m= 683.1500	Da	se(m)= 4.2 mDa	t=19.32 min	t=1159.3 s	se(t)=10.0 s	a= 253.8	Kcuentas	tiene 20 líneas
Comp.88	m= 581.1500	Da	se(m)= 2.1 mDa	t=19.42 min	t=1164.9 s	se(t)= 5.7 s	a= 228.2	Kcuentas	tiene 6 líneas
Comp.89	m= 579.1370	Da	se(m)= 0.9 mDa	t=19.47 min	t=1168.1 s	se(t)= 2.1 s	a= 619.1	Kcuentas	tiene 61 líneas
Comp.90	m= 579.1363	Da	se(m)= 0.8 mDa	t=19.64 min	t=1178.4 s	se(t)= 3.5 s	a= 925.7	Kcuentas	tiene 56 líneas
Comp.91	m= 597.2178	Da	se(m)= 5.5 mDa	t=19.87 min	t=1192.0 s	se(t)= 3.1 s	a= 805.6	Kcuentas	tiene 69 líneas
Comp.92	m= 449.1098	Da	se(m)= 1.7 mDa	t=20.02 min	t=1201.3 s	se(t)= 9.0 s	a= 236.5	Kcuentas	tiene 14 líneas
Comp.93	m= 579.1378	Da	se(m)= 1.1 mDa	t=20.13 min	t=1208.1 s	se(t)= 3.4 s	a= 2462.2	Kcuentas	tiene 52 líneas
Comp.94	m= 563.1448	Da	se(m)= 1.8 mDa	t=20.39 min	t=1223.1 s	se(t)= 2.9 s	a= 4072.8	Kcuentas	tiene 37 líneas
Comp.95	m= 641.2683	Da	se(m)= 1.0 mDa	t=20.47 min	t=1228.4 s	se(t)= 2.7 s	a= 536.6	Kcuentas	tiene 41 líneas
Comp.96	m= 341.0917	Da	se(m)= 0.6 mDa	t=20.66 min	t=1239.4 s	se(t)= 6.6 s	a= 1266.3	Kcuentas	tiene 28 líneas
Comp.97	m= 563.1420	Da	se(m)= 0.5 mDa	t=20.97 min	t=1258.2 s	se(t)= 4.5 s	a= 3772.4	Kcuentas	tiene 52 líneas
Comp.98	m= 593.1522	Da	se(m)= 1.1 mDa	t=21.09 min	t=1265.5 s	se(t)= 3.6 s	a= 734.1	Kcuentas	tiene 18 líneas
Comp.99	m= 363.0760	Da	se(m)= 0.8 mDa	t=21.16 min	t=1269.5 s	se(t)=12.5 s	a= 1346.4	Kcuentas	tiene 4 líneas
Comp.100	m= 447.0948	Da	se(m)= 0.8 mDa	t=21.18 min	t=1270.9 s	se(t)= 3.2 s	a= 2888.8	Kcuentas	tiene 38 líneas
Comp.101	m= 341.0915	Da	se(m)= 0.6 mDa	t=21.32 min	t=1279.2 s	se(t)=15.0 s	a= 4637.9	Kcuentas	tiene 7 líneas
Comp.102	m= 621.1487	Da	se(m)= 3.0 mDa	t=21.43 min	t=1285.9 s	se(t)= 4.2 s	a= 248.8	Kcuentas	tiene 52 líneas
Comp.103	m= 515.1247	Da	se(m)= 1.2 mDa	t=21.51 min	t=1290.9 s	se(t)=11.2 s	a= 705.0	Kcuentas	tiene 10 líneas
Comp.104	m= 641.2667	Da	se(m)= 3.0 mDa	t=21.58 min	t=1294.9 s	se(t)= 2.7 s	a= 423.4	Kcuentas	tiene 70 líneas
Comp.105	m= 530.1680	Da	se(m)= 2.1 mDa	t=21.72 min	t=1303.3 s	se(t)= 8.9 s	a= 258.4	Kcuentas	tiene 10 líneas
Comp.106	m= 447.0959	Da	se(m)= 0.7 mDa	t=21.84 min	t=1310.3 s	se(t)= 4.2 s	a= 3354.9	Kcuentas	tiene 54 líneas



Comp.107	m= 704.2230	Da	se (m)= 1.6 mDa	t=21.93 min	t=1315.5 s	se (t)= 3.9 s	a= 437.2	Kcuentas	tiene 6 líneas
Comp.108	m= 563.1449	Da	se (m)= 1.2 mDa	t=22.02 min	t=1321.5 s	se (t)= 3.1 s	a= 3807.3	Kcuentas	tiene 66 líneas
Comp.109	m= 491.2155	Da	se (m)= 0.9 mDa	t=22.17 min	t=1330.0 s	se (t)= 2.5 s	a= 363.4	Kcuentas	tiene 62 líneas
Comp.110	m= 405.1765	Da	se (m)= 4.9 mDa	t=22.31 min	t=1338.6 s	se (t)= 4.1 s	a= 396.7	Kcuentas	tiene 33 líneas
Comp.111	m= 593.1532	Da	se (m)= 2.1 mDa	t=22.42 min	t=1345.4 s	se (t)= 4.4 s	a= 541.1	Kcuentas	tiene 49 líneas
Comp.112	m= 577.1574	Da	se (m)= 1.2 mDa	t=22.51 min	t=1350.4 s	se (t)= 2.5 s	a= 282.8	Kcuentas	tiene 23 líneas
Comp.113	m= 363.0730	Da	se (m)= 0.5 mDa	t=22.61 min	t=1356.5 s	se (t)= 7.4 s	a= 3631.3	Kcuentas	tiene 39 líneas
Comp.114	m= 359.1488	Da	se (m)= 1.4 mDa	t=22.75 min	t=1364.7 s	se (t)= 3.1 s	a= 1284.9	Kcuentas	tiene 79 líneas
Comp.115	m= 433.1148	Da	se (m)= 1.3 mDa	t=22.89 min	t=1373.6 s	se (t)= 3.0 s	a= 332.3	Kcuentas	tiene 38 líneas
Comp.116	m= 609.1485	Da	se (m)= 1.4 mDa	t=23.08 min	t=1384.7 s	se (t)= 3.0 s	a= 290.3	Kcuentas	tiene 58 líneas
Comp.117	m= 695.1857	Da	se (m)= 1.1 mDa	t=23.21 min	t=1392.5 s	se (t)= 3.5 s	a= 401.7	Kcuentas	tiene 71 líneas
Comp.118	m= 363.0719	Da	se (m)= 1.0 mDa	t=23.45 min	t=1407.0 s	se (t)= 8.6 s	a= 664.3	Kcuentas	tiene 23 líneas
Comp.119	m= 433.1133	Da	se (m)= 1.0 mDa	t=23.58 min	t=1415.0 s	se (t)= 3.1 s	a= 493.2	Kcuentas	tiene 76 líneas
Comp.120	m= 361.1505	Da	se (m)= 0.5 mDa	t=23.75 min	t=1425.0 s	se (t)= 2.1 s	a= 350.8	Kcuentas	tiene 67 líneas
Comp.121	m= 199.0979	Da	se (m)= 1.0 mDa	t=23.79 min	t=1427.2 s	se (t)= 5.8 s	a= 298.1	Kcuentas	tiene 2 líneas
Comp.122	m= 579.1410	Da	se (m)= 2.1 mDa	t=23.92 min	t=1435.2 s	se (t)= 6.5 s	a= 11772.6	Kcuentas	tiene 33 líneas
Comp.123	m= 597.1841	Da	se (m)= 0.6 mDa	t=24.10 min	t=1446.0 s	se (t)= 2.8 s	a= 1840.4	Kcuentas	tiene 55 líneas
Comp.124	m= 461.1102	Da	se (m)= 0.8 mDa	t=24.27 min	t=1456.1 s	se (t)= 2.6 s	a= 223.2	Kcuentas	tiene 34 líneas
Comp.125	m= 463.0898	Da	se (m)= 0.9 mDa	t=24.37 min	t=1462.0 s	se (t)= 3.2 s	a= 1630.3	Kcuentas	tiene 53 líneas
Comp.126	m= 411.1332	Da	se (m)= 1.3 mDa	t=24.67 min	t=1480.3 s	se (t)= 7.6 s	a= 217.0	Kcuentas	tiene 17 líneas
Comp.127	m= 521.2032	Da	se (m)= 0.7 mDa	t=24.90 min	t=1493.9 s	se (t)= 3.7 s	a= 1111.3	Kcuentas	tiene 60 líneas
Comp.128	m= 533.2212	Da	se (m)= 4.7 mDa	t=25.01 min	t=1500.6 s	se (t)= 5.9 s	a= 247.9	Kcuentas	tiene 23 líneas
Comp.129	m= 343.1066	Da	se (m)= 0.8 mDa	t=25.29 min	t=1517.6 s	se (t)= 9.7 s	a= 794.9	Kcuentas	tiene 16 líneas
Comp.130	m= 593.1524	Da	se (m)= 1.3 mDa	t=25.42 min	t=1525.1 s	se (t)= 2.5 s	a= 255.2	Kcuentas	tiene 48 líneas
Comp.131	m= 621.1467	Da	se (m)= 1.0 mDa	t=25.56 min	t=1533.9 s	se (t)= 4.3 s	a= 425.9	Kcuentas	tiene 46 líneas
Comp.132	m= 199.0814	Da	se (m)= 1.5 mDa	t=25.70 min	t=1541.8 s	se (t)= 3.1 s	a= 863.5	Kcuentas	tiene 84 líneas
Comp.133	m= 623.1622	Da	se (m)= 1.4 mDa	t=25.80 min	t=1548.3 s	se (t)= 3.4 s	a= 333.2	Kcuentas	tiene 32 líneas
Comp.134	m= 535.2395	Da	se (m)= 1.7 mDa	t=25.96 min	t=1557.6 s	se (t)= 3.0 s	a= 271.6	Kcuentas	tiene 68 líneas
Comp.135	m= 343.1070	Da	se (m)= 0.9 mDa	t=26.06 min	t=1563.5 s	se (t)= 3.4 s	a= 231.4	Kcuentas	tiene 16 líneas
Comp.136	m= 565.1575	Da	se (m)= 0.8 mDa	t=26.19 min	t=1571.3 s	se (t)= 2.9 s	a= 1452.0	Kcuentas	tiene 64 líneas
Comp.137	m= 621.1475	Da	se (m)= 0.9 mDa	t=26.34 min	t=1580.6 s	se (t)= 4.7 s	a= 1630.0	Kcuentas	tiene 88 líneas
Comp.138	m= 567.1724	Da	se (m)= 0.7 mDa	t=26.44 min	t=1586.4 s	se (t)= 2.5 s	a= 590.6	Kcuentas	tiene 33 líneas
Comp.139	m= 447.1006	Da	se (m)= 2.6 mDa	t=26.63 min	t=1597.8 s	se (t)= 5.2 s	a= 14846.6	Kcuentas	tiene 35 líneas
Comp.140	m= 621.1508	Da	se (m)= 1.7 mDa	t=26.90 min	t=1613.9 s	se (t)= 5.7 s	a= 8805.8	Kcuentas	tiene 53 líneas
Comp.141	m= 675.3247	Da	se (m)= 0.9 mDa	t=27.04 min	t=1622.6 s	se (t)= 4.6 s	a= 595.5	Kcuentas	tiene 41 líneas
Comp.142	m= 447.0936	Da	se (m)= 1.2 mDa	t=27.13 min	t=1627.9 s	se (t)= 8.3 s	a= 391.1	Kcuentas	tiene 5 líneas
Comp.143	m= 675.3243	Da	se (m)= 0.6 mDa	t=27.23 min	t=1633.8 s	se (t)= 2.4 s	a= 480.3	Kcuentas	tiene 45 líneas
Comp.144	m= 513.2702	Da	se (m)= 0.8 mDa	t=27.36 min	t=1641.6 s	se (t)= 3.3 s	a= 484.1	Kcuentas	tiene 46 líneas
Comp.145	m= 187.0995	Da	se (m)= 0.6 mDa	t=27.46 min	t=1647.4 s	se (t)= 3.0 s	a= 5412.9	Kcuentas	tiene 73 líneas



Comp.146	m= 675.3240	Da	se(m)= 0.9	mDa	t=27.58 min	t=1655.0 s	se(t)= 4.3 s	a= 873.8	Kcuentas	tiene 70 líneas
Comp.147	m= 661.3089	Da	se(m)= 1.9	mDa	t=27.73 min	t=1663.6 s	se(t)= 6.0 s	a= 295.7	Kcuentas	tiene 27 líneas
Comp.148	m= 243.1243	Da	se(m)= 0.5	mDa	t=27.88 min	t=1673.0 s	se(t)= 3.2 s	a= 2349.9	Kcuentas	tiene 102 líneas
Comp.149	m= 513.2693	Da	se(m)= 1.6	mDa	t=27.89 min	t=1673.6 s	se(t)= 8.3 s	a= 349.8	Kcuentas	tiene 1 líneas
Comp.150	m= 245.1393	Da	se(m)= 1.5	mDa	t=28.04 min	t=1682.5 s	se(t)= 2.9 s	a= 519.4	Kcuentas	tiene 47 líneas
Comp.151	m= 411.1330	Da	se(m)= 1.3	mDa	t=28.06 min	t=1683.4 s	se(t)= 9.7 s	a= 408.0	Kcuentas	tiene 2 líneas
Comp.152	m= 472.1763	Da	se(m)= 1.0	mDa	t=28.19 min	t=1691.3 s	se(t)= 2.8 s	a= 549.7	Kcuentas	tiene 72 líneas
Comp.153	m= 621.3130	Da	se(m)= 2.7	mDa	t=28.46 min	t=1707.3 s	se(t)= 4.0 s	a= 232.0	Kcuentas	tiene 47 líneas
Comp.154	m= 675.3233	Da	se(m)= 1.7	mDa	t=28.57 min	t=1714.4 s	se(t)= 5.4 s	a= 409.4	Kcuentas	tiene 43 líneas
Comp.155	m= 761.3245	Da	se(m)= 1.1	mDa	t=28.71 min	t=1722.9 s	se(t)= 5.2 s	a= 591.9	Kcuentas	tiene 85 líneas
Comp.156	m= 539.1791	Da	se(m)= 1.0	mDa	t=28.89 min	t=1733.5 s	se(t)= 3.2 s	a= 3869.3	Kcuentas	tiene 54 líneas
Comp.157	m= 675.3235	Da	se(m)= 1.5	mDa	t=29.00 min	t=1740.0 s	se(t)= 3.4 s	a= 251.9	Kcuentas	tiene 37 líneas
Comp.158	m= 761.3249	Da	se(m)= 1.1	mDa	t=29.09 min	t=1745.5 s	se(t)= 6.7 s	a= 1985.2	Kcuentas	tiene 60 líneas
Comp.159	m= 675.3233	Da	se(m)= 1.5	mDa	t=29.22 min	t=1753.4 s	se(t)= 3.7 s	a= 289.6	Kcuentas	tiene 55 líneas
Comp.160	m= 331.1754	Da	se(m)= 0.8	mDa	t=29.35 min	t=1760.9 s	se(t)= 2.7 s	a= 230.0	Kcuentas	tiene 35 líneas
Comp.161	m= 461.1090	Da	se(m)= 0.6	mDa	t=29.45 min	t=1767.0 s	se(t)= 3.0 s	a= 1801.6	Kcuentas	tiene 85 líneas
Comp.162	m= 545.2606	Da	se(m)= 0.9	mDa	t=29.61 min	t=1776.4 s	se(t)= 3.6 s	a= 1389.3	Kcuentas	tiene 74 líneas
Comp.163	m= 199.0981	Da	se(m)= 0.5	mDa	t=29.74 min	t=1784.5 s	se(t)= 2.6 s	a= 224.8	Kcuentas	tiene 62 líneas
Comp.164	m= 433.1149	Da	se(m)= 0.5	mDa	t=29.92 min	t=1795.3 s	se(t)= 2.3 s	a= 2053.0	Kcuentas	tiene 103 líneas



Compuestos obtenidos mediante el *clustering* de líneas y 10 líneas de mayor área de la huella espectrométrica de cada uno. Se presentan únicamente aquellos que habían sido previamente identificados en la caracterización convencional de pimienta:

Comp 4	m=191.0589 Da	t=2.1 min	a=17124.0 Kcuentas
191.0589	2.1	17124.00000	100.0000
145.0629	2.1	1658.38400	9.6846
192.0603	2.1	1546.32000	9.0301
165.0415	2.1	739.54400	4.3188
371.1187	2.1	716.04800	4.1815
317.0544	2.1	694.63600	4.0565
325.1241	2.1	413.15200	2.4127
193.0607	2.1	359.13200	2.0972
293.0979	2.1	321.05600	1.8749
144.0672	2.2	221.12000	1.2913
Comp 12	m=191.0212 Da	t=3.4 min	a=4729.0 Kcuentas
191.0212	3.4	4728.98800	100.0000
Comp 31	m=243.0622 Da	t=6.6 min	a=1344.5 Kcuentas
243.0622	6.6	1344.54800	100.0000
292.1389	6.7	616.61200	45.8602
342.1175	6.6	480.27600	35.7203
200.0571	6.6	246.26800	18.3160
202.1079	6.7	221.20800	16.4522
362.0505	6.6	185.96000	13.8307
362.0497	6.7	184.00000	13.6849
244.0652	6.6	156.06800	11.6075
316.1502	6.5	151.47200	11.2656
260.0255	6.6	141.43600	10.5192
Comp 39	m=282.0836 Da	t=8.4 min	a=3710.3 Kcuentas
282.0836	8.4	3710.28400	100.0000
565.1760	8.5	2526.56400	68.0962
566.1779	8.5	672.32400	18.1206
283.0859	8.4	498.53200	13.4365
848.2676	8.5	447.49600	12.0610
587.1578	8.4	193.15600	5.2060
567.1807	8.5	142.64400	3.8446
550.1779	8.4	113.90400	3.0700
380.0581	8.4	106.91600	2.8816
663.1509	8.4	103.20400	2.7816
Comp 42	m=164.0730 Da	t=9.2 min	a=2410.0 Kcuentas
164.0730	9.2	2409.98800	100.0000
518.1821	9.2	347.58800	14.4228
165.0757	9.2	284.24400	11.7944
351.1317	9.2	274.02800	11.3705
201.1252	9.2	272.83200	11.3209
147.0461	9.2	215.18400	8.9288
405.1402	9.2	187.13600	7.7650
244.1295	9.1	164.90400	6.8425
570.2191	9.2	152.50000	6.3278
340.1246	9.3	116.30400	4.8259
Comp 44	m=299.0769 Da	t=9.4 min	a=243.5 Kcuentas
299.0769	9.4	243.52800	100.0000



288.1555	9.4	65.66400	26.9636	
315.0722	9.4	44.09600	18.1072	
272.1591	9.5	42.59200	17.4896	
616.2591	9.4	39.98400	16.4186	
424.1393	9.4	35.62800	14.6299	
300.0794	9.4	35.17200	14.4427	
457.1039	9.4	31.42400	12.9036	
137.0252	9.4	25.42000	10.4382	
504.1854	9.5	23.25600	9.5496	
Comp 46	m=326.1246	Da	t=9.9 min	a=1189.6 Kcuentas
326.1246	9.9	1189.64800	100.0000	
164.0726	9.9	915.33200	76.9414	
327.1239	9.8	191.52800	16.0996	
293.1130	9.9	102.33200	8.6019	
316.1736	9.9	82.92800	6.9708	
395.0704	9.9	43.46000	3.6532	
413.0870	9.9	27.42400	2.3052	
431.2093	9.8	14.22800	1.1960	
598.1753	9.9	13.60800	1.1439	
259.0652	9.8	12.89200	1.0837	
Comp 50	m=218.1041	Da	t=10.9 min	a=274.3 Kcuentas
218.1041	10.9	274.32000	100.0000	
566.1240	11.0	135.62800	49.4415	
396.1266	10.9	105.35600	38.4062	
344.1815	11.0	100.22000	36.5340	
481.1015	10.9	65.82000	23.9939	
642.0918	10.9	44.28400	16.1432	
472.2429	10.9	43.17600	15.7393	
422.1600	11.0	41.84800	15.2552	
373.1099	11.0	35.97600	13.1146	
514.2119	11.0	33.14800	12.0837	
Comp 52	m=329.0873	Da	t=11.3 min	a=457.5 Kcuentas
329.0873	11.3	457.54000	100.0000	
233.1287	11.3	272.96400	59.6590	
229.1560	11.4	125.70400	27.4739	
167.0373	11.4	118.53200	25.9064	
358.1969	11.3	99.19600	21.6803	
251.1027	11.4	77.76800	16.9970	
486.2564	11.4	72.98400	15.9514	
330.0915	11.3	67.12000	14.6698	
420.1433	11.4	59.32400	12.9659	
659.1832	11.4	42.33200	9.2521	
Comp 54	m=299.0773	Da	t=12.1 min	a=220.1 Kcuentas
299.0773	12.1	220.09600	100.0000	
460.0919	12.1	210.57600	95.6746	
390.0516	12.1	85.98800	39.0684	
437.2017	12.1	68.30800	31.0355	
325.0568	12.2	68.07600	30.9301	
235.1087	12.2	66.19600	30.0760	
265.0358	12.1	56.46400	25.6543	
548.2015	12.1	51.74000	23.5079	
292.1278	12.0	50.73200	23.0499	
261.0076	12.2	49.80800	22.6301	
Comp 57	m=203.0843	Da	t=12.8 min	a=2621.1 Kcuentas
203.0843	12.8	2621.08400	100.0000	
365.1351	12.8	353.78000	13.4975	
429.1554	12.8	350.90000	13.3876	



223.0258	12.9	322.69600	12.3115	
204.0865	12.8	296.12400	11.2978	
592.2372	12.9	257.48400	9.8236	
323.0876	12.8	226.04000	8.6239	
158.0837	12.8	166.14000	6.3386	
649.2097	12.8	110.88800	4.2306	
430.1593	12.8	97.34000	3.7137	
Comp 60	m=403.1254	Da	t=13.4 min	a=644.4 Kcuentas
403.1254	13.4	644.38400	100.0000	
329.0880	13.4	412.41600	64.0016	
503.1429	13.5	200.57200	31.1262	
193.0164	13.4	134.30400	20.8422	
404.1287	13.4	129.30000	20.0657	
425.1654	13.4	120.97200	18.7733	
330.0922	13.4	58.34800	9.0548	
504.1449	13.5	50.64800	7.8599	
299.0811	13.5	49.66800	7.7078	
720.2974	13.4	48.31200	7.4974	
Comp 61	m=343.1034	Da	t=13.5 min	a=453.1 Kcuentas
343.1034	13.5	453.12800	100.0000	
344.1069	13.5	78.40800	17.3037	
358.1981	13.5	48.78400	10.7661	
430.2307	13.5	47.38000	10.4562	
431.2180	13.5	29.91200	6.6012	
392.1499	13.5	17.88400	3.9468	
505.1477	13.5	15.92800	3.5151	
345.1098	13.5	15.42400	3.4039	
622.1497	13.5	14.40000	3.1779	
189.0774	13.5	12.70400	2.8036	
Comp 69	m=627.1579	Da	t=14.8 min	a=1011.0 Kcuentas
627.1579	14.8	1011.00800	100.0000	
628.1617	14.8	303.52400	30.0219	
759.2014	14.8	140.08000	13.8555	
629.1654	14.8	84.54000	8.3620	
649.1415	14.8	56.92400	5.6304	
427.1287	14.8	55.00800	5.4409	
495.2075	14.8	35.46800	3.5082	
487.1446	14.9	31.14400	3.0805	
299.0783	14.8	29.26000	2.8941	
435.1269	14.9	24.23200	2.3968	
Comp 72	m=597.1484	Da	t=16.1 min	a=2941.5 Kcuentas
597.1484	16.1	2941.45200	100.0000	
598.1511	16.1	865.36800	29.4198	
349.0611	16.2	421.69600	14.3363	
554.0988	16.1	141.39200	4.8069	
343.0945	16.2	69.93200	2.3775	
191.0581	16.1	41.18400	1.4001	
528.1583	16.2	31.88800	1.0841	
358.1969	16.2	29.28400	0.9956	
785.1577	16.1	8.44400	0.2871	
465.1981	16.1	6.44400	0.2191	
Comp 73	m=341.0887	Da	t=16.2 min	a=1811.3 Kcuentas
341.0887	16.2	1811.28800	100.0000	
179.0367	16.2	608.34800	33.5865	
526.1581	16.2	404.67600	22.3419	
342.0920	16.2	328.28400	18.1243	
683.1851	16.2	297.22800	16.4098	



705.1663	16.2	227.91200	12.5829	
408.1310	16.2	189.10800	10.4405	
527.1601	16.2	114.63200	6.3288	
684.1893	16.2	99.20000	5.4768	
706.1700	16.2	81.66400	4.5086	
Comp 78	m=609.1483 Da	t=17.7 min	a=2058.4 Kcuentas	
609.1483	17.7	2058.40400	100.0000	
406.1663	17.7	758.80000	36.8635	
610.1510	17.7	668.30400	32.4671	
611.1535	17.7	174.77200	8.4907	
177.0206	17.7	167.58000	8.1413	
407.1691	17.7	155.51200	7.5550	
167.0366	17.6	127.63200	6.2005	
363.0396	17.7	94.01200	4.5672	
241.0021	17.6	81.24800	3.9471	
345.0625	17.7	76.99600	3.7406	
Comp 80	m=581.1531 Da	t=18.0 min	a=10178.7 Kcuentas	
581.1531	18.0	10178.73200	100.0000	
582.1551	18.0	3510.88000	34.4923	
583.1574	18.0	894.06800	8.7837	
355.1023	17.9	397.93200	3.9094	
603.1336	18.0	300.06000	2.9479	
679.1222	18.0	186.36000	1.8309	
584.1599	18.0	164.07600	1.6119	
604.1375	18.0	114.68400	1.1267	
515.1237	18.1	108.22400	1.0632	
566.2926	18.0	105.94800	1.0409	
Comp 82	m=741.1893 Da	t=18.4 min	a=1751.4 Kcuentas	
741.1893	18.4	1751.42800	100.0000	
609.1470	18.4	1592.58800	90.9308	
742.1919	18.4	628.85600	35.9053	
610.1496	18.4	508.65600	29.0424	
445.2055	18.4	469.82400	26.8252	
385.1869	18.4	340.96000	19.4675	
743.1939	18.4	200.39200	11.4416	
387.1797	18.5	155.58800	8.8835	
370.0907	18.5	145.48400	8.3066	
447.0937	18.4	93.72400	5.3513	
Comp 85	m=579.1375 Da	t=19.0 min	a=2317.7 Kcuentas	
579.1375	19.0	2317.67200	100.0000	
580.1391	19.0	660.96000	28.5183	
581.1478	19.0	405.60800	17.5007	
281.1391	19.0	372.84800	16.0872	
415.1257	19.0	112.47200	4.8528	
582.1523	19.0	103.65600	4.4724	
408.1772	19.0	73.14400	3.1559	
565.1575	19.0	64.20400	2.7702	
282.1428	19.0	63.39600	2.7353	
363.1603	19.0	58.92000	2.5422	
Comp 86	m=593.1561 Da	t=19.1 min	a=5328.6 Kcuentas	
593.1561	19.1	5328.64400	100.0000	
594.1563	19.1	1817.68400	34.1116	
595.1559	19.1	442.64800	8.3070	
515.1243	19.1	136.96400	2.5703	
675.1557	19.1	94.28000	1.7693	
596.1586	19.1	84.80000	1.5914	
691.1234	19.1	71.08800	1.3341	



676.1591	19.1	31.23600	0.5862	
300.1913	19.1	28.22000	0.5296	
516.1283	19.1	27.38800	0.5140	
Comp 93	m=579.1378	Da	t=20.1 min	a=2462.2 Kcuentas
579.1378	20.1	2462.24400	100.0000	
595.1674	20.1	754.82800	30.6561	
580.1396	20.1	708.97200	28.7937	
581.1470	20.2	283.10400	11.4978	
596.1715	20.1	239.73600	9.7365	
307.1393	20.2	219.99200	8.9346	
677.1074	20.1	69.61200	2.8272	
582.1510	20.2	65.42800	2.6573	
429.1402	20.1	58.32000	2.3686	
661.1394	20.1	42.62800	1.7313	
Comp 94	m=563.1448	Da	t=20.4 min	a=4072.8 Kcuentas
563.1448	20.4	4072.80000	100.0000	
564.1454	20.4	1358.62800	33.3586	
565.1477	20.4	368.14000	9.0390	
363.0764	20.4	291.43200	7.1556	
623.1631	20.4	66.30800	1.6281	
364.0804	20.4	62.17200	1.5265	
645.1447	20.4	60.44000	1.4840	
661.1126	20.4	59.45600	1.4598	
566.1511	20.4	57.35600	1.4083	
415.1623	20.4	51.18000	1.2566	
Comp 100	m=447.0948	Da	t=21.2 min	a=2888.8 Kcuentas
447.0948	21.2	2888.81200	100.0000	
448.0975	21.2	702.28400	24.3105	
449.1010	21.2	131.70400	4.5591	
895.1975	21.2	64.56000	2.2348	
527.0520	21.2	57.92400	2.0051	
327.0514	21.2	35.11600	1.2156	
513.1105	21.1	34.49200	1.1940	
191.0571	21.3	33.23600	1.1505	
896.2000	21.2	31.55200	1.0922	
335.1711	21.1	26.06000	0.9021	
Comp 101	m=341.0915	Da	t=21.3 min	a=4637.9 Kcuentas
341.0915	21.3	4637.86800	100.0000	
342.0953	21.4	609.70400	13.1462	
543.1757	21.3	191.49600	4.1290	
365.0770	21.3	64.76000	1.3963	
544.1801	21.3	61.02800	1.3159	
311.1395	21.4	39.34000	0.8482	
528.0572	21.4	18.12800	0.3909	
Comp 106	m=447.0959	Da	t=21.8 min	a=3354.9 Kcuentas
447.0959	21.8	3354.93200	100.0000	
448.1002	21.8	749.72800	22.3470	
449.1020	21.8	137.68400	4.1039	
444.1459	21.9	114.66000	3.4177	
895.1991	21.8	105.37200	3.1408	
533.2247	21.8	100.52000	2.9962	
641.2679	21.8	90.82800	2.7073	
545.0673	21.8	85.08800	2.5362	
361.1346	21.8	83.12000	2.4775	
697.1683	21.8	62.00800	1.8483	
Comp 108	m=563.1449	Da	t=22.0 min	a=3807.3 Kcuentas
563.1449	22.0	3807.28400	100.0000	



564.1461	22.0	1273.58800	33.4514	
565.1489	22.0	350.61600	9.2091	
345.1564	22.0	129.26400	3.3952	
217.0341	22.0	99.20400	2.6056	
381.0851	22.0	91.72000	2.4091	
405.1789	22.1	70.02000	1.8391	
566.1514	22.0	60.22800	1.5819	
187.0991	22.1	58.51200	1.5368	
656.3253	22.0	50.60000	1.3290	
Comp 115	m=433.1148	Da	t=22.9 min	a=332.3 Kcuentas
433.1148	22.9	332.27200	100.0000	
434.1181	22.9	88.42400	26.6119	
691.3133	22.9	77.43200	23.3038	
641.2658	22.9	70.02000	21.0731	
551.2301	22.8	40.68400	12.2442	
340.1878	22.9	29.48000	8.8722	
349.1490	22.9	28.98000	8.7218	
553.2467	22.9	28.84800	8.6820	
607.1687	23.0	27.96400	8.4160	
692.3166	22.9	27.82400	8.3739	
Comp 119	m=433.1133	Da	t=23.6 min	a=493.2 Kcuentas
433.1133	23.6	493.22800	100.0000	
199.0257	23.6	298.54000	60.5278	
479.2126	23.5	232.10400	47.0582	
411.1290	23.6	183.53600	37.2112	
392.1821	23.6	163.60000	33.1692	
405.1771	23.6	129.00000	26.1542	
434.1171	23.6	124.20000	25.1811	
533.1278	23.6	113.59200	23.0303	
393.0837	23.6	84.97200	17.2277	
163.0406	23.6	82.74800	16.7768	
Comp 122	m=579.1410	Da	t=23.9 min	a=11772.6 Kcuentas
579.1410	23.9	11772.60000	100.0000	
580.1396	23.9	4074.35200	34.6088	
581.1410	23.9	1079.26000	9.1676	
677.1057	23.9	434.34400	3.6894	
582.1435	23.9	213.65200	1.8148	
447.0927	23.9	151.70000	1.2886	
615.1126	23.9	135.62400	1.1520	
601.1173	24.0	114.58000	0.9733	
285.0399	23.9	104.32000	0.8861	
678.1099	24.0	81.07200	0.6886	
Comp 123	m=597.1841	Da	t=24.1 min	a=1840.4 Kcuentas
597.1841	24.1	1840.37600	100.0000	
598.1866	24.1	588.96400	32.0024	
577.1569	24.2	549.48000	29.8569	
578.1598	24.2	175.17200	9.5183	
599.1895	24.1	161.96400	8.8006	
489.1035	24.0	159.80800	8.6834	
411.1318	24.1	85.88800	4.6669	
609.1482	24.1	69.53600	3.7784	
447.1868	24.1	61.05600	3.3176	
448.1671	24.2	46.51600	2.5275	
Comp 125	m=463.0898	Da	t=24.4 min	a=1630.3 Kcuentas
463.0898	24.4	1630.32000	100.0000	
447.0942	24.4	1486.98400	91.2081	
464.0929	24.4	391.27200	23.9997	



448.0978	24.4	359.96800	22.0796	
533.6161	24.3	93.43600	5.7311	
465.0958	24.4	82.71600	5.0736	
411.1319	24.3	77.52000	4.7549	
449.1000	24.4	66.75200	4.0944	
409.1168	24.4	43.98000	2.6976	
607.1654	24.4	43.27200	2.6542	
Comp 127	m=521.2032	Da	t=24.9 min	a=1111.3 Kcuentas
521.2032	24.9	1111.26800	100.0000	
522.2070	24.9	319.21200	28.7250	
461.1091	24.8	193.03600	17.3708	
448.1646	24.8	182.57200	16.4292	
603.2059	24.9	101.60000	9.1427	
433.2431	24.9	101.01200	9.0898	
523.2127	24.9	84.52000	7.6057	
305.1597	24.9	59.80000	5.3812	
449.1671	24.9	59.45600	5.3503	
427.1285	25.0	59.28400	5.3348	
Comp 132	m=199.0814	Da	t=25.7 min	a=863.5 Kcuentas
199.0814	25.7	863.45600	100.0000	
609.1467	25.8	318.27600	36.8607	
505.1001	25.7	286.61200	33.1936	
286.1111	25.7	220.91200	25.5846	
343.1069	25.7	184.47200	21.3644	
595.2113	25.7	168.87200	19.5577	
308.0934	25.7	143.29200	16.5952	
155.0911	25.7	140.70400	16.2954	
261.1340	25.6	139.07600	16.1069	
187.0993	25.6	130.94000	15.1646	
Comp 136	m=565.1575	Da	t=26.2 min	a=1452.0 Kcuentas
565.1575	26.2	1452.03600	100.0000	
563.1420	26.2	1008.00000	69.4198	
566.1600	26.2	444.01600	30.5789	
317.1616	26.2	360.27200	24.8115	
564.1449	26.2	292.66400	20.1554	
567.1628	26.2	114.53600	7.8880	
431.2282	26.2	100.28000	6.9062	
243.1240	26.1	89.78000	6.1830	
318.1650	26.1	59.65200	4.1082	
673.2389	26.2	58.08800	4.0005	
Comp 158	m=761.3249	Da	t=29.1 min	a=1985.2 Kcuentas
761.3249	29.1	1985.20400	100.0000	
431.0986	29.2	1296.06000	65.2860	
607.1676	29.1	1143.91600	57.6221	
762.3279	29.1	755.11200	38.0370	
555.2797	29.1	386.08000	19.4479	
717.3340	29.1	375.17200	18.8984	
608.1703	29.1	364.55200	18.3635	
432.1017	29.2	307.78000	15.5037	
763.3317	29.1	262.18400	13.2069	
651.1572	29.1	216.79600	10.9206	
Comp 161	m=461.1090	Da	t=29.5 min	a=1801.6 Kcuentas
461.1090	29.5	1801.64400	100.0000	
462.1116	29.4	409.39600	22.7235	
635.1611	29.5	374.54800	20.7892	
717.3327	29.5	257.51200	14.2932	
636.1654	29.5	131.05600	7.2742	



718.3370	29.5	118.90400	6.5998	
631.2945	29.4	110.00400	6.1058	
463.1142	29.4	92.83200	5.1526	
657.3094	29.5	86.52400	4.8025	
685.2349	29.5	76.40400	4.2408	
Comp 164	m=433.1149	Da	t=29.9 min	a=2053.0 Kcuentas
433.1149	29.9	2052.96800	100.0000	
434.1174	29.9	572.60400	27.8915	
706.2809	29.9	323.86800	15.7756	
513.2697	29.8	184.30800	8.9776	
271.0613	29.9	126.89200	6.1809	
707.2841	29.9	123.72800	6.0268	
435.1198	29.9	114.04000	5.5549	
673.3070	29.9	100.54000	4.8973	
867.2381	29.9	85.34400	4.1571	
649.2567	30.0	70.37600	3.4280	



Comparativa de áreas de los compuestos obtenidos con el *clustering* de líneas entre las tres muestras de pimiento:

Se presenta la masa exacta, el tiempo, el área y el área relativa.

EMPAREJAMOS MUESTRAS

Area>100.000

Muestra v1: 11.342 clusters

Muestra a1: 14.624 clusters

Muestra r1: 13.321 clusters

Numero de clusters por encima del umbral de area: 6432

COMPUESTO 1

MUESTRA 1

191.0589	2.1	17124.0	100.00
192.0603	2.1	1546.3	100.00
193.0607	2.1	359.1	100.00
194.0636	2.0	11.6	100.00
194.0644	2.2	12.6	100.00

MUESTRA 2

191.0561	2.2	8081.4	47.19
192.0593	2.2	621.5	40.19
195.0514	2.2	1166.7	324.87
196.0551	2.2	74.8	643.77

MUESTRA 3

191.0578	2.2	13410.4	78.31
192.0604	2.2	1067.3	69.02
193.0609	2.2	179.3	49.92
194.0651	2.1	11.7	100.34
196.0564	2.2	22.3	177.70

COMPUESTO 2

MUESTRA 1

447.1006	26.6	14846.6	100.00
448.0987	26.6	4818.2	100.00
449.1000	26.6	969.9	100.00
450.1029	26.6	154.9	100.00
451.1044	26.6	11.9	100.00
451.1073	26.7	10.0	100.00

MUESTRA 2

447.0965	26.6	8442.4	56.86
448.0966	26.6	2536.0	52.63
449.0984	26.6	516.8	53.28
450.1005	26.6	78.1	50.44
451.1069	26.6	11.8	99.19

MUESTRA 3

447.0967	26.7	7291.7	49.11
448.0975	26.6	2060.0	42.76
449.0993	26.6	412.3	42.51
450.1023	26.6	66.1	42.70
451.1079	26.6	6.1	51.20

COMPUESTO 3

MUESTRA 1

579.1410	23.9	11772.6	100.00
----------	------	---------	--------



580.1396	23.9	4074.4	100.00
581.1410	23.9	1079.3	100.00
582.1435	23.9	213.7	100.00
583.1483	23.9	27.3	100.00
MUESTRA 2			
579.1373	23.9	5292.5	44.96
580.1380	23.9	1680.7	41.25
581.1396	23.9	425.2	39.40
582.1437	23.9	85.2	39.89
583.1464	23.9	8.8	32.15
MUESTRA 3			
579.1364	23.9	3420.6	29.06
580.1378	23.9	960.6	23.58
581.1402	23.9	257.6	23.87
582.1468	23.9	44.7	20.90
COMPUESTO 4			
MUESTRA 1			
581.1531	18.0	10178.7	100.00
582.1551	18.0	3510.9	100.00
583.1574	18.0	894.1	100.00
584.1599	18.0	164.1	100.00
585.1609	17.9	25.5	100.00
MUESTRA 2			
581.1486	18.0	590.3	5.80
582.1513	18.0	173.6	4.94
583.1472	18.1	11.6	1.29
583.1500	17.9	41.9	25.56
MUESTRA 3			
581.1534	17.9	113.5	1.12
582.1571	17.9	19.7	0.56
583.1525	17.9	6.7	0.75
COMPUESTO 5			
MUESTRA 1			
621.1508	26.9	8805.8	100.00
622.1518	26.9	3095.8	100.00
623.1539	26.9	799.4	100.00
624.1558	26.9	145.5	100.00
625.1592	26.9	28.6	100.00
MUESTRA 2			
621.1464	26.9	5855.7	66.50
622.1478	26.9	1928.5	62.29
623.1502	26.9	513.1	64.18
624.1524	26.9	100.4	69.00
625.1580	26.9	14.7	51.59
MUESTRA 3			
621.1487	26.9	3860.6	43.84
622.1514	26.9	1294.4	41.81
623.1531	26.9	344.0	43.03
624.1558	26.9	57.6	39.57
COMPUESTO 6			
MUESTRA 1			
665.1408	26.9	7296.4	100.00
666.1417	26.9	2664.4	100.00
667.1429	26.9	771.4	100.00
668.1461	26.9	151.2	100.00



669.1499	26.9	17.1	100.00
669.1509	26.8	12.8	100.00
MUESTRA 2			
665.1363	26.9	5336.2	73.14
666.1382	26.9	1904.9	71.49
667.1402	26.9	517.1	67.04
668.1435	26.9	101.8	67.36
669.1477	26.9	16.2	94.94
MUESTRA 3			
665.1390	26.9	3407.0	46.69
666.1411	26.9	1144.8	42.97
667.1441	26.9	315.8	40.93
668.1467	26.9	65.6	43.37

COMPUESTO 7

MUESTRA 1			
761.3249	29.1	1985.2	31.09
762.3279	29.1	755.1	32.89
763.3317	29.1	262.2	37.41
764.3359	29.1	68.2	40.85
MUESTRA 2			
761.3229	29.0	6385.6	100.00
762.3245	29.1	2296.1	100.00
763.3269	29.1	700.9	100.00
764.3314	29.1	167.0	100.00
765.3335	29.1	26.4	100.00
MUESTRA 3			
761.3247	29.1	423.3	6.63
762.3276	29.1	170.3	7.42
763.3292	29.1	64.6	9.22
764.3354	29.1	9.4	5.61

COMPUESTO 8

MUESTRA 1			
187.0995	27.5	5412.9	100.00
188.1022	27.5	588.4	100.00
189.1034	27.5	69.8	100.00
MUESTRA 2			
187.0976	27.5	126.5	2.34
188.1001	27.4	6.7	1.13
MUESTRA 3			
187.0981	27.5	229.2	4.23
188.1010	27.5	24.6	4.17

COMPUESTO 9

MUESTRA 1			
593.1561	19.1	5328.6	100.00
594.1563	19.1	1817.7	100.00
595.1559	19.1	442.6	100.00
596.1586	19.1	84.8	100.00
597.1620	19.1	14.8	100.00
MUESTRA 2			
593.1486	19.1	1385.7	26.00
594.1523	19.1	410.7	22.59
595.1506	19.1	108.0	24.40
MUESTRA 3			
593.1516	19.1	773.5	14.52
594.1563	19.1	218.4	12.02



595.1559	19.1	57.6	13.01
COMPUESTO 10			
MUESTRA 1			
000.0000	00.0	0000.0	000.00
MUESTRA 2			
191.0203	3.1	4400.6	90.51
192.0232	3.0	367.1	134.57
193.0247	3.2	30.9	69.54
193.0252	3.0	44.2	0.00
MUESTRA 3			
191.0207	3.0	4862.2	100.00
192.0239	3.0	272.8	100.00
193.0248	3.0	44.4	100.00
COMPUESTO 11			
MUESTRA 1			
191.0212	3.4	4729.0	100.00
192.0244	3.5	201.9	100.00
193.0261	3.3	19.3	100.00
193.0265	3.4	15.6	100.00
MUESTRA 2			
191.0200	3.4	2867.7	60.64
192.0230	3.3	157.0	77.78
193.0244	3.3	44.6	231.84
MUESTRA 3			
191.0208	3.4	3012.4	63.70
192.0238	3.4	286.0	141.70
193.0254			



APÉNDICE 2: *CISTUS*.

Se ha elegido una muestra representante del grupo de *Cistus* con las siguientes características:

Nombre de la muestra	S4
intervalo masas analizadas	[0 Da, 1200 Da]
intervalo de tiempo analizado	[0 min, 38 min]
Tamaño fichero	3 MB

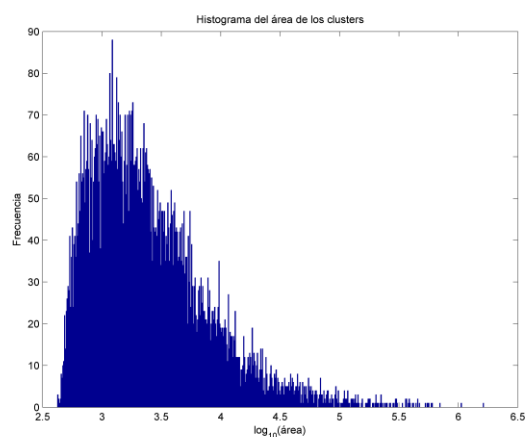
Se ha realizado la transformación del formato baf a XML, después se ha convertido en formato binario, se ha eliminado el ruido y se ha calibrado en masa. Se ha realizado el *clustering* de líneas utilizando los siguientes parámetros:

Nombre del parámetro	Valor utilizado
umbral_l	100 cuentas
umbral_area	500
T_resol_fase1	300s
T_resol_fase2	20s
umbral_tr	20s
umbral_m/z	0.010Da



Tablas con el número de *clusters* y el tiempo computacional, y el histograma del área de los *clusters* obtenidos con el *clustering* de líneas:

Clustering de líneas	Nº de clusters
Número inicial de puntos (tr,m/z,l) sin ruido	391.262
Número de puntos (tr,m/z,l) sin ruido, umbral_l=100	353.630
<i>Pre-Clustering</i> fase1: T_resol=300s	9.718
<i>Pre-Clustering</i> fase2: T_resol=20s	20.017
Elimina: puntos<3, area<500	8.137
<i>Clustering</i> de puntos	30.789
Elimina: puntos<3, area<500	11.099
Detección de picos cromatogr.	21.516
Elimina: puntos<3, area<500	18.695
Umbral de area> 1000	13.852
Umbral de area> 10 ⁴	2.004
Umbral de area> 10 ⁵	195
Umbral de area> 10 ⁶	14



Tiempo computacional

<i>Precluster.:</i> fase 1 y 2	2.4 min	37.5 %
<i>Clustering</i> de puntos	1.6 min	25 %
Detección p. cromatogr.	1.5 min	23.5 %
Estadística	0.9 min	14 %
TOTAL	6.4 min	100%



Lista de compuestos obtenidos con el *clustering* de líneas para la muestra representativa de *Cistus*:

Muestra: 1. Fichero con clusters: cistus11.m
 Umbral area: 200000 cuentas
 Numero de clusters iniciales: 2941

Lista de compuestos ordenados por tiempo de retención

Comp.1	m= 715.0387	Da	se(m)= 0.4	mDa	t= 0.44	min	t= 26.6	s	se(t)= 4.8	s	a= 8565.0	Kcuentas	tiene	96	líneas
Comp.2	m= 232.9753	Da	se(m)= 1.0	mDa	t= 0.76	min	t= 45.5	s	se(t)= 8.8	s	a= 131.8	Kcuentas	tiene	2	líneas
Comp.3	m= 201.0250	Da	se(m)= 0.2	mDa	t= 1.75	min	t=105.3	s	se(t)= 3.3	s	a= 619.1	Kcuentas	tiene	177	líneas
Comp.4	m= 317.0519	Da	se(m)= 1.1	mDa	t= 1.96	min	t=117.6	s	se(t)= 3.1	s	a= 75.1	Kcuentas	tiene	12	líneas
Comp.5	m= 209.0324	Da	se(m)= 7.2	mDa	t= 2.06	min	t=123.9	s	se(t)= 5.6	s	a= 896.4	Kcuentas	tiene	29	líneas
Comp.6	m= 355.1134	Da	se(m)=10.0	mDa	t= 2.16	min	t=129.7	s	se(t)= 3.6	s	a= 315.7	Kcuentas	tiene	25	líneas
Comp.7	m= 191.0550	Da	se(m)= 0.5	mDa	t= 2.29	min	t=137.5	s	se(t)= 4.0	s	a= 8822.4	Kcuentas	tiene	45	líneas
Comp.8	m= 355.1225	Da	se(m)= 0.7	mDa	t= 2.40	min	t=143.9	s	se(t)= 4.2	s	a= 285.7	Kcuentas	tiene	6	líneas
Comp.9	m= 649.1781	Da	se(m)= 1.0	mDa	t= 2.55	min	t=152.8	s	se(t)= 4.5	s	a= 292.9	Kcuentas	tiene	8	líneas
Comp.10	m= 355.1224	Da	se(m)= 0.5	mDa	t= 2.64	min	t=158.7	s	se(t)= 4.2	s	a= 433.3	Kcuentas	tiene	13	líneas
Comp.11	m= 311.0965	Da	se(m)= 1.0	mDa	t= 2.86	min	t=171.8	s	se(t)= 5.4	s	a= 112.5	Kcuentas	tiene	6	líneas
Comp.12	m= 173.0444	Da	se(m)= 0.3	mDa	t= 3.00	min	t=180.2	s	se(t)= 3.4	s	a= 1518.4	Kcuentas	tiene	33	líneas
Comp.13	m= 420.9111	Da	se(m)= 1.2	mDa	t= 3.28	min	t=196.8	s	se(t)= 6.0	s	a= 105.7	Kcuentas	tiene	11	líneas
Comp.14	m= 297.1164	Da	se(m)= 0.7	mDa	t= 3.62	min	t=217.2	s	se(t)= 4.3	s	a= 317.9	Kcuentas	tiene	5	líneas
Comp.15	m= 297.1173	Da	se(m)= 1.1	mDa	t= 4.00	min	t=240.2	s	se(t)= 4.0	s	a= 127.9	Kcuentas	tiene	5	líneas
Comp.16	m= 337.0753	Da	se(m)= 0.5	mDa	t= 4.11	min	t=246.6	s	se(t)= 4.4	s	a= 852.5	Kcuentas	tiene	8	líneas
Comp.17	m= 481.0604	Da	se(m)= 0.9	mDa	t= 4.23	min	t=253.8	s	se(t)= 5.1	s	a= 531.9	Kcuentas	tiene	5	líneas
Comp.18	m= 436.9635	Da	se(m)= 0.5	mDa	t= 4.52	min	t=270.9	s	se(t)= 4.6	s	a= 235.5	Kcuentas	tiene	17	líneas
Comp.19	m= 481.0582	Da	se(m)= 1.5	mDa	t= 4.75	min	t=284.9	s	se(t)= 5.2	s	a= 91.3	Kcuentas	tiene	4	líneas
Comp.20	m= 481.0598	Da	se(m)= 1.2	mDa	t= 5.05	min	t=302.9	s	se(t)= 6.2	s	a= 182.7	Kcuentas	tiene	4	líneas
Comp.21	m= 481.0593	Da	se(m)= 0.6	mDa	t= 6.64	min	t=398.3	s	se(t)= 5.2	s	a= 439.3	Kcuentas	tiene	4	líneas
Comp.22	m= 271.0796	Da	se(m)= 0.4	mDa	t= 7.31	min	t=438.5	s	se(t)= 3.5	s	a= 162.5	Kcuentas	tiene	7	líneas
Comp.23	m= 611.1384	Da	se(m)= 1.9	mDa	t= 7.66	min	t=459.3	s	se(t)= 4.2	s	a= 73.0	Kcuentas	tiene	4	líneas
Comp.24	m= 737.2314	Da	se(m)= 0.9	mDa	t= 7.86	min	t=471.4	s	se(t)= 3.3	s	a= 199.1	Kcuentas	tiene	7	líneas
Comp.25	m= 685.1769	Da	se(m)= 1.4	mDa	t= 7.97	min	t=478.1	s	se(t)= 6.6	s	a= 148.6	Kcuentas	tiene	2	líneas
Comp.26	m= 331.0650	Da	se(m)= 0.7	mDa	t= 8.12	min	t=487.1	s	se(t)= 3.7	s	a= 199.7	Kcuentas	tiene	3	líneas
Comp.27	m= 647.1988	Da	se(m)= 0.5	mDa	t= 8.33	min	t=499.8	s	se(t)= 3.5	s	a= 254.1	Kcuentas	tiene	5	líneas
Comp.28	m= 169.0127	Da	se(m)= 0.2	mDa	t= 8.52	min	t=511.1	s	se(t)= 4.0	s	a= 963.6	Kcuentas	tiene	12	líneas
Comp.29	m= 633.0677	Da	se(m)= 1.5	mDa	t= 8.75	min	t=524.8	s	se(t)= 3.9	s	a= 78.1	Kcuentas	tiene	2	líneas
Comp.30	m= 331.0639	Da	se(m)= 0.7	mDa	t= 8.89	min	t=533.4	s	se(t)= 3.6	s	a= 257.5	Kcuentas	tiene	7	líneas
Comp.31	m= 609.1172	Da	se(m)= 0.9	mDa	t= 9.14	min	t=548.2	s	se(t)= 3.7	s	a= 101.0	Kcuentas	tiene	6	líneas



Comp.32	m= 779.2400	Da	se(m)= 0.9 mDa	t= 9.32 min	t=559.2 s	se(t)= 3.7 s	a= 280.4	Kcuentas	tiene	8 líneas
Comp.33	m= 483.0743	Da	se(m)= 1.4 mDa	t= 9.60 min	t=576.2 s	se(t)= 3.5 s	a= 57.7	Kcuentas	tiene	4 líneas
Comp.34	m= 331.0646	Da	se(m)= 1.4 mDa	t=10.04 min	t=602.2 s	se(t)= 4.7 s	a= 176.9	Kcuentas	tiene	5 líneas
Comp.35	m= 782.0501	Da	se(m)= 1.8 mDa	t=10.21 min	t=612.3 s	se(t)= 2.1 s	a= 62.8	Kcuentas	tiene	6 líneas
Comp.36	m= 781.0490	Da	se(m)= 1.0 mDa	t=10.32 min	t=619.2 s	se(t)= 6.7 s	a= 461.1	Kcuentas	tiene	7 líneas
Comp.37	m= 411.0205	Da	se(m)= 1.5 mDa	t=10.54 min	t=632.2 s	se(t)= 4.2 s	a= 99.3	Kcuentas	tiene	7 líneas
Comp.38	m= 609.1187	Da	se(m)= 1.5 mDa	t=10.76 min	t=645.9 s	se(t)= 5.2 s	a= 96.3	Kcuentas	tiene	5 líneas
Comp.39	m= 783.0666	Da	se(m)= 1.9 mDa	t=10.87 min	t=651.9 s	se(t)= 3.6 s	a= 69.4	Kcuentas	tiene	3 líneas
Comp.40	m= 315.0693	Da	se(m)= 0.7 mDa	t=10.98 min	t=659.0 s	se(t)= 3.9 s	a= 239.0	Kcuentas	tiene	8 líneas
Comp.41	m= 345.0800	Da	se(m)= 1.6 mDa	t=11.24 min	t=674.3 s	se(t)= 3.4 s	a= 51.8	Kcuentas	tiene	6 líneas
Comp.42	m= 467.1160	Da	se(m)= 0.7 mDa	t=11.51 min	t=690.6 s	se(t)= 3.3 s	a= 83.7	Kcuentas	tiene	9 líneas
Comp.43	m= 541.0216	Da	se(m)= 0.9 mDa	t=11.62 min	t=697.3 s	se(t)= 6.4 s	a= 235.0	Kcuentas	tiene	9 líneas
Comp.44	m= 609.1190	Da	se(m)= 0.6 mDa	t=11.72 min	t=703.2 s	se(t)= 4.7 s	a= 1722.0	Kcuentas	tiene	17 líneas
Comp.45	m= 607.0508	Da	se(m)=10.9 mDa	t=11.89 min	t=713.7 s	se(t)= 8.1 s	a= 228.5	Kcuentas	tiene	5 líneas
Comp.46	m= 541.5239	Da	se(m)= 1.3 mDa	t=12.03 min	t=721.7 s	se(t)= 5.2 s	a= 203.6	Kcuentas	tiene	11 líneas
Comp.47	m= 301.0528	Da	se(m)= 2.7 mDa	t=12.13 min	t=727.9 s	se(t)= 5.9 s	a= 700.9	Kcuentas	tiene	13 líneas
Comp.48	m= 541.5234	Da	se(m)= 0.9 mDa	t=12.25 min	t=735.1 s	se(t)= 2.5 s	a= 67.9	Kcuentas	tiene	13 líneas
Comp.49	m= 783.0618	Da	se(m)= 2.1 mDa	t=12.40 min	t=743.8 s	se(t)= 5.5 s	a= 136.2	Kcuentas	tiene	6 líneas
Comp.50	m= 305.0634	Da	se(m)= 0.5 mDa	t=12.70 min	t=761.8 s	se(t)= 4.9 s	a= 288.8	Kcuentas	tiene	19 líneas
Comp.51	m= 365.0160	Da	se(m)= 1.2 mDa	t=12.95 min	t=777.2 s	se(t)= 8.2 s	a= 220.0	Kcuentas	tiene	5 líneas
Comp.52	m= 603.0290	Da	se(m)= 1.8 mDa	t=13.06 min	t=783.4 s	se(t)= 5.9 s	a= 80.2	Kcuentas	tiene	5 líneas
Comp.53	m= 301.0899	Da	se(m)= 0.4 mDa	t=13.17 min	t=790.1 s	se(t)= 3.8 s	a= 1056.7	Kcuentas	tiene	32 líneas
Comp.54	m= 727.1868	Da	se(m)= 1.4 mDa	t=13.27 min	t=796.0 s	se(t)= 4.5 s	a= 209.3	Kcuentas	tiene	6 líneas
Comp.55	m= 359.0952	Da	se(m)= 0.6 mDa	t=13.36 min	t=801.7 s	se(t)= 3.3 s	a= 99.5	Kcuentas	tiene	5 líneas
Comp.56	m= 313.0893	Da	se(m)= 0.8 mDa	t=13.57 min	t=814.3 s	se(t)= 3.5 s	a= 60.5	Kcuentas	tiene	12 líneas
Comp.57	m= 608.1068	Da	se(m)= 4.0 mDa	t=13.76 min	t=825.6 s	se(t)= 8.0 s	a= 85.7	Kcuentas	tiene	6 líneas
Comp.58	m= 760.1231	Da	se(m)= 2.9 mDa	t=13.85 min	t=831.0 s	se(t)= 6.7 s	a= 110.3	Kcuentas	tiene	8 líneas
Comp.59	m= 285.0583	Da	se(m)= 0.4 mDa	t=13.94 min	t=836.4 s	se(t)= 4.7 s	a= 1783.1	Kcuentas	tiene	33 líneas
Comp.60	m= 433.1314	Da	se(m)= 0.5 mDa	t=14.04 min	t=842.6 s	se(t)= 3.8 s	a= 277.3	Kcuentas	tiene	9 líneas
Comp.61	m= 913.1733	Da	se(m)= 1.2 mDa	t=14.18 min	t=850.8 s	se(t)= 3.9 s	a= 212.0	Kcuentas	tiene	17 líneas
Comp.62	m= 761.1280	Da	se(m)= 1.1 mDa	t=14.34 min	t=860.4 s	se(t)= 6.1 s	a= 983.2	Kcuentas	tiene	18 líneas
Comp.63	m=1085.0521	Da	se(m)= 2.0 mDa	t=14.51 min	t=870.8 s	se(t)= 3.5 s	a= 61.4	Kcuentas	tiene	25 líneas
Comp.64	m= 541.0218	Da	se(m)= 0.8 mDa	t=14.57 min	t=874.4 s	se(t)= 9.2 s	a= 4294.2	Kcuentas	tiene	21 líneas
Comp.65	m= 593.1255	Da	se(m)= 1.3 mDa	t=14.69 min	t=881.5 s	se(t)= 3.4 s	a= 62.1	Kcuentas	tiene	4 líneas
Comp.66	m= 355.0627	Da	se(m)= 1.2 mDa	t=14.78 min	t=886.7 s	se(t)= 5.6 s	a= 63.1	Kcuentas	tiene	5 líneas
Comp.67	m= 359.1319	Da	se(m)= 1.0 mDa	t=14.95 min	t=896.8 s	se(t)= 3.1 s	a= 51.3	Kcuentas	tiene	8 líneas
Comp.68	m= 407.0974	Da	se(m)= 0.9 mDa	t=15.04 min	t=902.3 s	se(t)= 5.4 s	a= 203.2	Kcuentas	tiene	23 líneas
Comp.69	m= 603.0272	Da	se(m)= 1.5 mDa	t=15.16 min	t=909.8 s	se(t)=10.1 s	a= 1916.2	Kcuentas	tiene	24 líneas
Comp.70	m= 542.0268	Da	se(m)= 2.0 mDa	t=15.28 min	t=916.9 s	se(t)= 6.9 s	a= 204.2	Kcuentas	tiene	14 líneas



Comp.71	m= 608.0940	Da	se(m)= 1.8 mDa	t=15.46 min	t=927.5 s	se(t)= 7.2 s	a= 97.0	Kcuentas tiene 16 líneas
Comp.72	m= 541.0200	Da	se(m)= 1.2 mDa	t=15.66 min	t=939.5 s	se(t)= 4.1 s	a= 153.2	Kcuentas tiene 20 líneas
Comp.73	m= 541.5226	Da	se(m)= 1.4 mDa	t=15.75 min	t=945.1 s	se(t)= 3.4 s	a= 65.4	Kcuentas tiene 9 líneas
Comp.74	m= 603.0264	Da	se(m)= 1.1 mDa	t=15.90 min	t=953.7 s	se(t)= 5.5 s	a= 270.0	Kcuentas tiene 26 líneas
Comp.75	m= 293.1204	Da	se(m)= 0.7 mDa	t=16.03 min	t=961.6 s	se(t)= 3.7 s	a= 243.0	Kcuentas tiene 10 líneas
Comp.76	m= 305.0630	Da	se(m)= 0.6 mDa	t=16.15 min	t=969.2 s	se(t)= 3.9 s	a= 457.8	Kcuentas tiene 22 líneas
Comp.77	m= 603.0272	Da	se(m)= 0.6 mDa	t=16.34 min	t=980.4 s	se(t)= 7.8 s	a= 2541.3	Kcuentas tiene 27 líneas
Comp.78	m=1084.0468	Da	se(m)= 1.3 mDa	t=16.50 min	t=989.8 s	se(t)= 4.4 s	a= 262.3	Kcuentas tiene 30 líneas
Comp.79	m= 541.0208	Da	se(m)= 0.4 mDa	t=16.56 min	t=993.5 s	se(t)=12.1 s	a= 5933.5	Kcuentas tiene 16 líneas
Comp.80	m= 605.1421	Da	se(m)= 2.0 mDa	t=16.78 min	t=1006.6 s	se(t)= 3.6 s	a= 57.1	Kcuentas tiene 11 líneas
Comp.81	m= 542.0278	Da	se(m)= 0.8 mDa	t=17.11 min	t=1026.3 s	se(t)= 7.2 s	a= 558.8	Kcuentas tiene 15 líneas
Comp.82	m= 627.1844	Da	se(m)= 1.5 mDa	t=17.33 min	t=1039.5 s	se(t)= 3.7 s	a= 111.6	Kcuentas tiene 11 líneas
Comp.83	m= 588.0220	Da	se(m)= 1.1 mDa	t=17.43 min	t=1045.6 s	se(t)= 4.8 s	a= 155.2	Kcuentas tiene 12 líneas
Comp.84	m= 181.0846	Da	se(m)= 0.4 mDa	t=17.62 min	t=1056.9 s	se(t)= 3.7 s	a= 145.4	Kcuentas tiene 24 líneas
Comp.85	m= 401.1425	Da	se(m)= 2.1 mDa	t=17.71 min	t=1062.9 s	se(t)=3.4 s	a= 66.9	Kcuentas tiene 4 líneas
Comp.86	m= 603.0277	Da	se(m)= 1.9 mDa	t=17.80 min	t=1067.9 s	se(t)= 6.3 s	a= 311.1	Kcuentas tiene 22 líneas
Comp.87	m= 165.0901	Da	se(m)= 1.0 mDa	t=17.92 min	t=1075.1 s	se(t)= 3.7 s	a= 99.4	Kcuentas tiene 10 líneas
Comp.88	m= 595.1599	Da	se(m)= 1.2 mDa	t=18.25 min	t=1094.8 s	se(t)= 4.6 s	a= 429.1	Kcuentas tiene 20 líneas
Comp.89	m= 542.0274	Da	se(m)= 3.1 mDa	t=18.40 min	t=1103.8 s	se(t)=10.9 s	a= 119.9	Kcuentas tiene 3 líneas
Comp.90	m= 327.1057	Da	se(m)= 1.3 mDa	t=18.50 min	t=1110.1 s	se(t)= 3.8 s	a= 119.2	Kcuentas tiene 13 líneas
Comp.91	m= 757.1530	Da	se(m)= 2.1 mDa	t=18.59 min	t=1115.1 s	se(t)= 3.9 s	a= 84.0	Kcuentas tiene 8 líneas
Comp.92	m= 541.0206	Da	se(m)= 2.8 mDa	t=18.69 min	t=1121.7 s	se(t)= 4.2 s	a= 50.5	Kcuentas tiene 9 líneas
Comp.93	m= 163.0746	Da	se(m)= 0.8 mDa	t=18.88 min	t=1132.8 s	se(t)= 3.7 s	a= 112.5	Kcuentas tiene 18 líneas
Comp.94	m= 181.0854	Da	se(m)= 0.3 mDa	t=18.97 min	t=1138.0 s	se(t)= 4.5 s	a= 694.9	Kcuentas tiene 25 líneas
Comp.95	m= 385.1832	Da	se(m)= 0.9 mDa	t=19.16 min	t=1149.8 s	se(t)= 3.8 s	a= 179.0	Kcuentas tiene 19 líneas
Comp.96	m= 489.1910	Da	se(m)= 1.0 mDa	t=19.28 min	t=1157.1 s	se(t)= 3.6 s	a= 127.2	Kcuentas tiene 10 líneas
Comp.97	m= 179.0700	Da	se(m)= 0.3 mDa	t=19.44 min	t=1166.3 s	se(t)= 4.4 s	a= 754.9	Kcuentas tiene 20 líneas
Comp.98	m= 599.2258	Da	se(m)= 1.3 mDa	t=19.57 min	t=1173.9 s	se(t)= 4.0 s	a= 83.3	Kcuentas tiene 10 líneas
Comp.99	m= 588.0208	Da	se(m)= 1.2 mDa	t=19.83 min	t=1189.6 s	se(t)= 8.7 s	a= 210.0	Kcuentas tiene 7 líneas
Comp.100	m= 453.1680	Da	se(m)= 4.0 mDa	t=20.03 min	t=1201.8 s	se(t)= 1.9 s	a= 94.7	Kcuentas tiene10 líneas
Comp.101	m= 245.0463	Da	se(m)= 0.3 mDa	t=20.20 min	t=1212.3 s	se(t)= 8.4 s	a= 2185.5	Kcuentas tiene 7 líneas
Comp.102	m= 457.0730	Da	se(m)= 0.7 mDa	t=20.53 min	t=1231.8 s	se(t)= 4.7 s	a= 393.0	Kcuentas tiene18 líneas
Comp.103	m= 447.1461	Da	se(m)= 1.1 mDa	t=20.63 min	t=1238.0 s	se(t)= 3.3 s	a= 59.8	Kcuentas tiene10 líneas
Comp.104	m= 181.0863	Da	se(m)= 0.6 mDa	t=20.85 min	t=1250.8 s	se(t)= 3.8 s	a= 84.4	Kcuentas tiene14 líneas
Comp.105	m= 327.1422	Da	se(m)= 0.7 mDa	t=20.94 min	t=1256.4 s	se(t)= 4.7 s	a= 8114.1	Kcuentas tiene31 líneas
Comp.106	m= 631.0867	Da	se(m)= 0.7 mDa	t=21.09 min	t=1265.5 s	se(t)= 5.6 s	a= 1569.0	Kcuentas tiene 9 líneas
Comp.107	m= 489.1890	Da	se(m)= 1.3 mDa	t=21.23 min	t=1274.1 s	se(t)= 3.6 s	a= 129.6	Kcuentas tiene 11líneas
Comp.108	m= 357.1515	Da	se(m)= 1.2 mDa	t=21.68 min	t=1300.7 s	se(t)= 4.1 s	a= 61.1	Kcuentas tiene 6 líneas
Comp.109	m= 457.1106	Da	se(m)= 0.6 mDa	t=21.81 min	t=1308.4 s	se(t)= 7.1 s	a= 1895.1	Kcuentas tiene 8 líneas



Comp.110	m= 611.2465	Da	se (m)= 1.0	mDa	t=21.99	min	t=1319.6	s	se (t)= 6.1	s	a= 325.2	Kcuentas	tiene	16	líneas
Comp.111	m= 385.1093	Da	se (m)= 0.9	mDa	t=22.16	min	t=1329.8	s	se (t)= 3.9	s	a= 85.5	Kcuentas	tiene	7	líneas
Comp.112	m= 479.0771	Da	se (m)= 0.4	mDa	t=22.26	min	t=1335.6	s	se (t)= 3.5	s	a= 3237.0	Kcuentas	tiene	18	líneas
Comp.113	m= 495.1434	Da	se (m)= 1.8	mDa	t=22.34	min	t=1340.7	s	se (t)= 3.4	s	a= 59.4	Kcuentas	tiene	5	líneas
Comp.114	m= 479.0778	Da	se (m)= 0.6	mDa	t=22.52	min	t=1351.4	s	se (t)= 6.0	s	a= 1418.7	Kcuentas	tiene	14	líneas
Comp.115	m= 495.1458	Da	se (m)= 0.9	mDa	t=22.75	min	t=1365.0	s	se (t)= 5.0	s	a= 241.7	Kcuentas	tiene	9	líneas
Comp.116	m= 595.1224	Da	se (m)= 1.4	mDa	t=22.91	min	t=1374.4	s	se (t)= 5.2	s	a= 297.7	Kcuentas	tiene	14	líneas
Comp.117	m= 449.0683	Da	se (m)= 2.1	mDa	t=23.13	min	t=1387.9	s	se (t)= 4.0	s	a= 222.6	Kcuentas	tiene	10	líneas
Comp.118	m= 615.0914	Da	se (m)= 0.6	mDa	t=23.32	min	t=1398.9	s	se (t)= 4.6	s	a= 1003.9	Kcuentas	tiene	13	líneas
Comp.119	m= 433.0382	Da	se (m)= 1.7	mDa	t=23.67	min	t=1420.2	s	se (t)= 5.7	s	a= 1062.9	Kcuentas	tiene	10	líneas
Comp.120	m= 615.0913	Da	se (m)= 1.8	mDa	t=23.84	min	t=1430.3	s	se (t)= 5.3	s	a= 174.4	Kcuentas	tiene	6	líneas
Comp.121	m= 449.0674	Da	se (m)= 0.3	mDa	t=23.97	min	t=1437.9	s	se (t)= 4.2	s	a= 1377.6	Kcuentas	tiene	24	líneas
Comp.122	m= 565.2209	Da	se (m)= 0.9	mDa	t=24.13	min	t=1447.7	s	se (t)= 4.4	s	a= 211.2	Kcuentas	tiene	9	líneas
Comp.123	m= 433.1088	Da	se (m)= 2.1	mDa	t=24.26	min	t=1455.7	s	se (t)= 3.5	s	a= 115.9	Kcuentas	tiene	9	líneas
Comp.124	m= 449.0684	Da	se (m)= 0.6	mDa	t=24.40	min	t=1463.9	s	se (t)=5.1	s	a= 1680.1	Kcuentas	tiene	19	líneas
Comp.125	m= 487.1228	Da	se (m)= 1.8	mDa	t=24.48	min	t=1468.9	s	se (t)= 6.5	s	a= 165.7	Kcuentas	tiene	4	líneas
Comp.126	m= 463.0828	Da	se (m)= 0.3	mDa	t=24.70	min	t=1482.0	s	se (t)= 4.1	s	a= 2774.5	Kcuentas	tiene	27	líneas
Comp.127	m= 505.2016	Da	se (m)= 0.6	mDa	t=24.88	min	t=1492.7	s	se (t)= 3.9	s	a= 360.8	Kcuentas	tiene	19	líneas
Comp.128	m= 463.0827	Da	se (m)= 0.6	mDa	t=25.02	min	t=1501.4	s	se (t)= 5.8	s	a= 1068.9	Kcuentas	tiene	12	líneas
Comp.129	m= 165.0895	Da	se (m)= 0.5	mDa	t=25.14	min	t=1508.3	s	se (t)= 4.4	s	a= 241.6	Kcuentas	tiene	9	líneas
Comp.130	m= 523.2104	Da	se (m)= 0.6	mDa	t=25.64	min	t=1538.1	s	se (t)= 2.7	s	a= 168.5	Kcuentas	tiene	9	líneas
Comp.131	m= 433.0718	Da	se (m)= 0.8	mDa	t=25.89	min	t=1553.6	s	se (t)= 4.4	s	a= 371.9	Kcuentas	tiene	11	líneas
Comp.132	m= 487.1225	Da	se (m)= 1.3	mDa	t=26.18	min	t=1570.7	s	se (t)= 7.0	s	a= 88.0	Kcuentas	tiene	2	líneas
Comp.133	m= 523.2116	Da	se (m)= 0.7	mDa	t=26.27	min	t=1576.3	s	se (t)= 3.8	s	a= 231.5	Kcuentas	tiene	12	líneas
Comp.134	m= 433.0728	Da	se (m)= 0.4	mDa	t=26.36	min	t=1581.5	s	se (t)= 4.8	s	a= 1474.0	Kcuentas	tiene	13	líneas
Comp.135	m= 617.1149	Da	se (m)= 2.5	mDa	t=26.47	min	t=1588.5	s	se (t)= 6.4	s	a= 92.2	Kcuentas	tiene	6	líneas
Comp.136	m= 479.1492	Da	se (m)= 0.9	mDa	t=26.59	min	t=1595.4	s	se (t)= 5.6	s	a= 148.4	Kcuentas	tiene	3	líneas
Comp.137	m= 341.1343	Da	se (m)= 1.8	mDa	t=26.74	min	t=1604.1	s	se (t)= 3.5	s	a= 60.3	Kcuentas	tiene	4	líneas
Comp.138	m= 433.0710	Da	se (m)= 0.4	mDa	t=26.97	min	t=1618.0	s	se (t)= 5.3	s	a= 2185.7	Kcuentas	tiene	21	líneas
Comp.139	m= 447.0863	Da	se (m)= 1.2	mDa	t=27.12	min	t=1627.1	s	se (t)= 6.9	s	a= 611.2	Kcuentas	tiene	6	líneas
Comp.140	m= 565.2195	Da	se (m)= 0.6	mDa	t=27.33	min	t=1639.9	s	se (t)= 4.0	s	a= 350.2	Kcuentas	tiene	19	líneas
Comp.141	m= 493.1998	Da	se (m)= 1.3	mDa	t=27.47	min	t=1647.9	s	se (t)= 3.6	s	a= 80.4	Kcuentas	tiene	7	líneas
Comp.142	m= 439.0998	Da	se (m)= 0.6	mDa	t=27.67	min	t=1660.2	s	se (t)= 5.7	s	a= 406.3	Kcuentas	tiene	12	líneas
Comp.143	m= 609.1151	Da	se (m)= 1.6	mDa	t=27.82	min	t=1669.2	s	se (t)=10.2	s	a= 222.3	Kcuentas	tiene	5	líneas
Comp.144	m= 537.1531	Da	se (m)= 1.3	mDa	t=27.85	min	t=1671.3	s	se (t)= 4.0	s	a= 119.1	Kcuentas	tiene	7	líneas
Comp.145	m= 417.0758	Da	se (m)= 0.9	mDa	t=27.99	min	t=1679.1	s	se (t)= 4.0	s	a= 147.7	Kcuentas	tiene	12	líneas
Comp.146	m= 537.1525	Da	se (m)= 0.4	mDa	t=28.29	min	t=1697.2	s	se (t)= 4.0	s	a= 255.3	Kcuentas	tiene	17	líneas
Comp.147	m= 415.1184	Da	se (m)= 1.6	mDa	t=28.41	min	t=1704.8	s	se (t)= 8.1	s	a= 55.6	Kcuentas	tiene	2	líneas
Comp.148	m= 441.1147	Da	se (m)= 1.1	mDa	t=28.59	min	t=1715.4	s	se (t)= 4.8	s	a= 65.9	Kcuentas	tiene	4	líneas



Comp.149	m= 289.1071	Da	se (m)= 1.3	mDa	t=28.74	min	t=1724.6	s	se (t)= 5.3	s	a= 72.2	Kcuentas	tiene	5	líneas
Comp.150	m= 441.1147	Da	se (m)= 1.2	mDa	t=28.85	min	t=1730.9	s	se (t)= 4.4	s	a= 55.8	Kcuentas	tiene	4	líneas
Comp.151	m= 601.1167	Da	se (m)= 3.4	mDa	t=29.04	min	t=1742.7	s	se (t)= 7.7	s	a= 66.0	Kcuentas	tiene	4	líneas
Comp.152	m= 625.1072	Da	se (m)= 1.0	mDa	t=29.14	min	t=1748.5	s	se (t)= 4.6	s	a= 100.7	Kcuentas	tiene	6	líneas
Comp.153	m= 435.1121	Da	se (m)=11.4	mDa	t=29.38	min	t=1762.9	s	se (t)= 5.1	s	a= 288.2	Kcuentas	tiene	8	líneas
Comp.154	m= 435.0878	Da	se (m)= 0.8	mDa	t=29.62	min	t=1776.9	s	se (t)= 1.7	s	a= 57.2	Kcuentas	tiene	10	líneas
Comp.155	m= 431.1856	Da	se (m)= 1.4	mDa	t=29.75	min	t=1785.3	s	se (t)= 3.5	s	a= 73.4	Kcuentas	tiene	3	líneas
Comp.156	m= 405.1130	Da	se (m)= 0.6	mDa	t=29.87	min	t=1792.3	s	se (t)= 4.4	s	a= 890.0	Kcuentas	tiene	15	líneas
Comp.157	m= 329.1358	Da	se (m)= 1.2	mDa	t=30.79	min	t=1847.4	s	se (t)= 3.2	s	a= 50.7	Kcuentas	tiene	7	líneas
Comp.158	m= 419.0939	Da	se (m)= 0.8	mDa	t=30.96	min	t=1857.6	s	se (t)= 5.1	s	a= 538.6	Kcuentas	tiene	8	líneas
Comp.159	m= 609.1162	Da	se (m)= 0.8	mDa	t=31.63	min	t=1897.6	s	se (t)= 4.3	s	a= 129.6	Kcuentas	tiene	5	líneas
Comp.160	m= 232.9732	Da	se (m)= 0.7	mDa	t=31.85	min	t=1911.0	s	se (t)= 5.6	s	a= 59.4	Kcuentas	tiene	2	líneas
Comp.161	m= 609.1150	Da	se (m)= 1.2	mDa	t=32.05	min	t=1922.9	s	se (t)= 7.2	s	a= 456.0	Kcuentas	tiene	4	líneas
Comp.162	m= 232.9732	Da	se (m)= 1.4	mDa	t=32.20	min	t=1932.1	s	se (t)= 5.9	s	a= 58.4	Kcuentas	tiene	2	líneas
Comp.163	m= 585.2244	Da	se (m)= 1.1	mDa	t=32.63	min	t=1957.9	s	se (t)=3.7	s	a= 98.2	Kcuentas	tiene	5	líneas
Comp.164	m= 429.0979	Da	se (m)= 1.0	mDa	t=32.82	min	t=1969.5	s	se (t)= 3.7	s	a= 76.2	Kcuentas	tiene	5	líneas
Comp.165	m= 232.9723	Da	se (m)= 0.9	mDa	t=33.15	min	t=1989.1	s	se (t)= 9.3	s	a= 112.8	Kcuentas	tiene	2	líneas
Comp.166	m= 537.1517	Da	se (m)= 1.3	mDa	t=33.51	min	t=2010.4	s	se (t)= 4.1	s	a= 185.9	Kcuentas	tiene	11	líneas
Comp.167	m= 521.1599	Da	se (m)= 1.4	mDa	t=33.60	min	t=2015.8	s	se (t)= 4.0	s	a= 131.9	Kcuentas	tiene	3	líneas
Comp.168	m= 232.9702	Da	se (m)= 0.7	mDa	t=33.80	min	t=2027.9	s	se (t)= 4.1	s	a= 51.1	Kcuentas	tiene	1	líneas
Comp.169	m= 593.1189	Da	se (m)= 1.0	mDa	t=33.98	min	t=2038.9	s	se (t)= 4.6	s	a= 893.7	Kcuentas	tiene	6	líneas
Comp.170	m= 232.9719	Da	se (m)= 1.3	mDa	t=34.20	min	t=2051.8	s	se (t)= 8.2	s	a= 108.9	Kcuentas	tiene	1	líneas
Comp.171	m= 593.1186	Da	se (m)= 1.1	mDa	t=34.72	min	t=2083.2	s	se (t)= 5.2	s	a= 332.5	Kcuentas	tiene	3	líneas
Comp.172	m= 232.9723	Da	se (m)= 1.1	mDa	t=34.81	min	t=2088.5	s	se (t)= 7.6	s	a= 97.7	Kcuentas	tiene	2	líneas
Comp.173	m= 315.1771	Da	se (m)= 0.8	mDa	t=35.26	min	t=2115.5	s	se (t)= 3.8	s	a= 165.7	Kcuentas	tiene	7	líneas
Comp.174	m= 232.9733	Da	se (m)= 1.2	mDa	t=35.42	min	t=2125.4	s	se (t)= 5.7	s	a= 73.7	Kcuentas	tiene	3	líneas
Comp.175	m= 311.1640	Da	se (m)= 1.3	mDa	t=35.61	min	t=2136.8	s	se (t)= 6.7	s	a= 80.6	Kcuentas	tiene	2	líneas
Comp.176	m= 311.1646	Da	se (m)= 1.0	mDa	t=36.08	min	t=2165.0	s	se (t)= 4.3	s	a= 56.0	Kcuentas	tiene	2	líneas
Comp.177	m= 311.1647	Da	se (m)= 0.9	mDa	t=36.48	min	t=2188.7	s	se (t)= 4.2	s	a= 65.1	Kcuentas	tiene	2	líneas
Comp.178	m= 327.2117	Da	se (m)= 1.5	mDa	t=36.69	min	t=2201.3	s	se (t)= 7.6	s	a= 115.0	Kcuentas	tiene	4	líneas
Comp.179	m= 311.1632	Da	se (m)= 1.0	mDa	t=36.92	min	t=2215.4	s	se (t)= 9.9	s	a= 195.7	Kcuentas	tiene	3	líneas
Comp.180	m= 243.0619	Da	se (m)= 1.2	mDa	t=37.12	min	t=2227.2	s	se (t)= 3.3	s	a= 53.6	Kcuentas	tiene	5	líneas
Comp.181	m= 311.1639	Da	se (m)= 0.6	mDa	t=37.32	min	t=2239.3	s	se (t)= 4.3	s	a= 95.7	Kcuentas	tiene	7	líneas
Comp.182	m= 311.1643	Da	se (m)= 0.5	mDa	t=37.52	min	t=2251.1	s	se (t)= 2.5	s	a= 60.1	Kcuentas	tiene	1	líneas
Comp.183	m= 232.9711	Da	se (m)= 0.7	mDa	t=37.59	min	t=2255.2	s	se (t)= 6.8	s	a= 105.5	Kcuentas	tiene	4	líneas
Comp.184	m= 311.1643	Da	se (m)= 0.7	mDa	t=37.85	min	t=2270.7	s	se (t)= 5.1	s	a= 128.3	Kcuentas	tiene	5	líneas



Compuestos obtenidos mediante el *clustering* de líneas y 10 líneas de mayor área de la huella espectrométrica de cada uno. Se presentan únicamente aquellos que habían sido previamente identificados en la caracterización convencional de *Cistus*:

Comp 7	m=191.0550 Da	t=2.3 min	a=8822.4 Kcuentas
191.0550	2.3	8822.37800	100.0000
192.0583	2.3	626.10900	7.0968
487.1267	2.2	499.92500	5.6666
193.0636	2.3	200.60700	2.2738
649.1788	2.2	138.14700	1.5659
179.0545	2.2	136.07800	1.5424
267.0699	2.3	111.61600	1.2651
488.1307	2.2	97.64400	1.1068
207.0489	2.3	77.52100	0.8787
189.0395	2.3	65.87600	0.7467
Comp 12	m=173.0444 Da	t=3.0 min	a=1518.4 Kcuentas
173.0444	3.0	1518.35800	100.0000
643.1682	2.9	301.08900	19.8299
529.1750	3.0	271.18600	17.8605
575.0678	3.0	130.84000	8.6172
174.0482	3.0	113.48700	7.4743
591.1733	3.0	64.99700	4.2807
207.0501	3.0	64.94500	4.2773
839.2650	3.0	61.88100	4.0755
355.1234	3.0	61.68400	4.0625
267.0704	2.9	61.16700	4.0285
Comp 16	m=337.0753 Da	t=4.1 min	a=852.5 Kcuentas
337.0753	4.1	852.51400	100.0000
605.1897	4.1	261.45100	30.6682
338.0787	4.1	115.62800	13.5632
606.1944	4.1	60.58900	7.1071
277.0547	4.1	24.30700	2.8512
339.0809	4.1	23.32400	2.7359
607.1921	4.1	13.51500	1.5853
390.9976	4.1	10.16000	1.1918
Comp 17	m=481.0604 Da	t=4.2 min	a=531.9 Kcuentas
481.0604	4.2	531.90400	100.0000
482.0632	4.2	120.02600	22.5654
297.1161	4.3	64.37300	12.1024
483.0652	4.2	23.87700	4.4890
480.0465	4.2	10.10500	1.8998
Comp 19	m=481.0582 Da	t=4.7 min	a=91.3 Kcuentas
481.0582	4.7	91.28400	100.0000
343.0670	4.7	31.36800	34.3631
482.0621	4.8	15.20100	16.6524
335.0974	4.7	11.70100	12.8182
Comp 20	m=481.0598 Da	t=5.0 min	a=182.7 Kcuentas
481.0598	5.0	182.71200	100.0000
723.2153	5.0	40.11300	21.9542
482.0631	5.0	34.43000	18.8439
335.0948	5.1	22.24500	12.1749
Comp 21	m=481.0593 Da	t=6.6 min	a=439.3 Kcuentas
481.0593	6.6	439.26700	100.0000
482.0608	6.6	97.24700	22.1385
483.0637	6.6	19.88700	4.5273
479.0430	6.6	6.75800	1.5385



Comp 22	m=271.0796	Da	t=7.3 min	a=162.5	Kcuentas
271.0796	7.3	162.48100	100.0000		
633.1838	7.3	23.94700	14.7383		
385.0205	7.3	22.97600	14.1407		
272.0847	7.3	18.90500	11.6352		
281.1182	7.3	15.89500	9.7827		
366.1367	7.2	10.52700	6.4789		
295.1016	7.4	9.82700	6.0481		
Comp 26	m=331.0650	Da	t=8.1 min	a=199.7	Kcuentas
331.0650	8.1	199.69100	100.0000		
332.0670	8.1	29.65600	14.8509		
445.0049	8.1	7.10000	3.5555		
Comp 28	m=169.0127	Da	t=8.5 min	a=963.6	Kcuentas
169.0127	8.5	963.56100	100.0000		
391.9440	8.5	189.96600	19.7150		
125.0247	8.5	87.60000	9.0913		
170.0163	8.5	64.36800	6.6802		
392.9469	8.5	35.78500	3.7138		
390.9399	8.5	25.45000	2.6412		
282.9510	8.5	19.72000	2.0466		
763.2065	8.5	14.72500	1.5282		
343.0404	8.6	13.85000	1.4374		
685.1763	8.5	12.06500	1.2521		
Comp 30	m=331.0639	Da	t=8.9 min	a=257.5	Kcuentas
331.0639	8.9	257.51800	100.0000		
332.0671	8.9	31.11300	12.0819		
645.1807	9.0	25.71500	9.9857		
169.0124	8.9	24.16700	9.3846		
445.0015	8.9	11.13900	4.3255		
232.9728	8.9	9.44300	3.6669		
309.1158	8.9	6.04100	2.3459		
Comp 31	m=609.1172	Da	t=9.1 min	a=101.0	Kcuentas
609.1172	9.1	100.96700	100.0000		
777.2220	9.2	46.96400	46.5142		
610.1212	9.1	32.97100	32.6552		
647.1994	9.1	21.16200	20.9593		
778.2298	9.1	11.57100	11.4602		
232.9746	9.2	9.22400	9.1357		
Comp 34	m=331.0646	Da	t=10.0 min	a=176.9	Kcuentas
331.0646	10.0	176.93200	100.0000		
301.0893	10.1	50.73200	28.6732		
332.0656	10.0	24.93800	14.0947		
345.0802	10.1	22.09400	12.4873		
232.9742	10.1	6.78800	3.8365		
Comp 35	m=782.0501	Da	t=10.2 min	a=62.8	Kcuentas
782.0501	10.2	62.75900	100.0000		
299.0737	10.2	43.76700	69.7382		
777.2255	10.2	35.23900	56.1497		
600.9849	10.2	26.53800	42.2856		
601.9894	10.2	9.80800	15.6280		
309.1178	10.3	6.28000	10.0065		
Comp 36	m=781.0490	Da	t=10.3 min	a=461.1	Kcuentas
781.0490	10.3	461.12000	100.0000		
782.0512	10.4	111.51000	24.1824		
783.0534	10.3	53.61500	11.6271		
390.0219	10.3	47.75600	10.3565		
600.9850	10.4	31.32400	6.7930		
299.9879	10.3	8.02600	1.7405		
232.9729	10.4	5.57500	1.2090		



Comp 38	m=609.1187	Da	t=10.8 min	a=96.3 Kcuentas
609.1187	10.8	96.27900	100.0000	
610.1209	10.8	30.91100	32.1057	
913.1747	10.8	22.40500	23.2709	
232.9764	10.8	12.45700	12.9384	
607.0981	10.8	11.05100	11.4781	
Comp 44	m=609.1190	Da	t=11.7 min	a=1722.0 Kcuentas
609.1190	11.7	1721.95700	100.0000	
610.1228	11.7	569.59300	33.0782	
611.1249	11.7	135.76100	7.8841	
541.5234	11.6	134.15300	7.7907	
441.0787	11.7	105.03400	6.0997	
532.0183	11.7	91.66800	5.3235	
607.5481	11.8	61.89400	3.5944	
532.5207	11.7	50.66000	2.9420	
401.1056	11.7	45.20400	2.6252	
439.0554	11.7	33.07600	1.9208	
Comp 50	m=305.0634	Da	t=12.7 min	a=288.8 Kcuentas
305.0634	12.7	288.78000	100.0000	
603.0277	12.7	159.23100	55.1392	
765.1595	12.7	117.23900	40.5980	
603.5287	12.8	113.45600	39.2880	
541.0200	12.7	87.60900	30.3376	
391.0293	12.7	80.06300	27.7246	
766.1584	12.7	46.34500	16.0485	
306.0664	12.7	43.31500	14.9993	
541.5213	12.8	36.73500	12.7208	
669.5513	12.6	34.75300	12.0344	
Comp 59	m=285.0583	Da	t=13.9 min	a=1783.1 Kcuentas
285.0583	13.9	1783.13900	100.0000	
286.0614	13.9	257.04500	14.4153	
152.0095	13.9	78.17600	4.3842	
719.1963	14.0	74.45900	4.1757	
571.1233	13.9	56.65700	3.1774	
1065.1792	13.9	50.24800	2.8180	
287.0632	13.9	48.74700	2.7338	
625.0410	13.9	39.92700	2.2391	
683.1041	13.9	37.00800	2.0754	
169.0121	13.9	36.40200	2.0415	
Comp 62	m=761.1280	Da	t=14.3 min	a=983.2 Kcuentas
761.1280	14.3	983.18100	100.0000	
762.1319	14.3	394.71300	40.1465	
469.0002	14.3	304.21500	30.9419	
684.1129	14.4	110.62800	11.2520	
763.1324	14.3	108.79300	11.0654	
425.0101	14.3	83.46000	8.4888	
325.0899	14.4	75.82400	7.7121	
470.0040	14.3	63.87700	6.4970	
163.0378	14.4	41.60800	4.2320	
760.1242	14.3	38.44500	3.9103	
Comp 64	m=541.0218	Da	t=14.6 min	a=4294.2 Kcuentas
541.0218	14.6	4294.16200	100.0000	
541.5234	14.6	2298.74600	53.5319	
542.0244	14.6	941.86500	21.9336	
1083.0447	14.5	326.81000	7.6106	
542.5265	14.6	266.97100	6.2171	
1084.0480	14.5	175.62700	4.0899	
540.5178	14.6	123.38200	2.8732	
567.4764	14.6	86.31600	2.0101	



543.0251	14.6	70.95500	1.6524	
760.1286	14.6	59.27400	1.3803	
Comp 69	m=603.0272	Da	t=15.2 min	a=1916.2 Kcuentas
603.0272	15.2	1916.22500	100.0000	
603.5288	15.2	1169.88300	61.0514	
541.0202	15.2	499.71800	26.0783	
604.0297	15.2	444.03800	23.1725	
625.0215	15.1	344.38600	17.9721	
541.5214	15.1	265.37600	13.8489	
625.5230	15.1	202.34600	10.5596	
783.0606	15.2	195.37500	10.1958	
540.5159	15.1	140.67000	7.3410	
604.5302	15.1	139.43800	7.2767	
Comp 79	m=541.0208	Da	t=16.6 min	a=5933.5 Kcuentas
541.0208	16.6	5933.52900	100.0000	
541.5228	16.6	3280.18500	55.2822	
542.0239	16.6	1205.97900	20.3248	
1083.0427	16.5	507.50700	8.5532	
542.5256	16.6	384.52200	6.4805	
540.5174	16.5	178.94500	3.0158	
543.0259	16.6	99.29900	1.6735	
567.4750	16.6	65.34100	1.1012	
300.9946	16.5	57.44400	0.9681	
567.9791	16.6	45.06600	0.7595	
423.0862	16.7	10.95500	19.2008	
Comp 81	m=542.0278	Da	t=17.1 min	a=558.8 Kcuentas
542.0278	17.1	558.77800	100.0000	
542.5296	17.1	308.77400	55.2588	
163.0380	17.1	123.55300	22.1113	
543.0323	17.1	113.98600	20.3992	
603.0284	17.1	83.43000	14.9308	
335.0771	17.1	65.30800	11.6876	
325.0889	17.1	56.99000	10.1990	
759.1171	17.2	48.07800	8.6041	
417.1021	17.1	33.45900	5.9879	
596.0205	17.1	31.92500	5.7134	
Comp 83	m=588.0220	Da	t=17.4 min	a=155.2 Kcuentas
588.0220	17.4	155.17200	100.0000	
289.0685	17.4	145.26500	93.6155	
541.0205	17.5	99.16200	63.9046	
588.5241	17.4	89.67300	57.7894	
603.0277	17.4	57.15600	36.8340	
589.0275	17.4	26.43200	17.0340	
290.0730	17.3	17.99600	11.5975	
759.1121	17.5	14.69000	9.4669	
771.1923	17.5	13.71600	8.8392	
604.0319	17.4	6.56900	4.2334	
Comp 86	m=603.0277	Da	t=17.8 min	a=311.1 Kcuentas
603.0277	17.8	311.11800	100.0000	
588.0223	17.8	222.93800	71.6571	
603.5270	17.8	197.86400	63.5977	
541.0204	17.8	114.89300	36.9291	
588.5234	17.8	100.61200	32.3389	
329.0838	17.8	96.22200	30.9278	
625.0239	17.8	79.25500	25.4743	
604.0304	17.8	78.91000	25.3634	
625.5254	17.8	47.74800	15.3472	
541.5222	17.8	41.45900	13.3258	
Comp 88	m=595.1599	Da	t=18.2 min	a=429.1 Kcuentas



595.1599	18.2	429.14600	100.0000	
596.1646	18.2	128.22000	29.8779	
602.0195	18.2	94.10200	21.9277	
603.0281	18.3	90.30300	21.0425	
373.1461	18.3	49.70500	11.5823	
607.0930	18.2	48.91400	11.3980	
211.0968	18.3	48.04900	11.1964	
602.5224	18.2	43.64200	10.1695	
603.5313	18.3	40.30500	9.3919	
401.1412	18.3	39.38800	9.1782	
Comp 90	m=327.1057	Da	t=18.5 min	a=119.2 Kcuentas
327.1057	18.5	119.22000	100.0000	
469.1080	18.5	102.94800	86.3513	
541.5215	18.5	53.76200	45.0948	
541.0218	18.5	53.55600	44.9220	
371.0952	18.5	47.14300	39.5429	
481.0921	18.5	40.36900	33.8609	
489.1922	18.5	24.76700	20.7742	
760.1214	18.5	21.62400	18.1379	
470.1140	18.5	18.25400	15.3112	
328.1100	18.5	17.53200	14.7056	
Comp 91	m=757.1530	Da	t=18.6 min	a=84.0 Kcuentas
757.1530	18.6	83.96200	100.0000	
179.0698	18.6	74.16100	88.3269	
341.1191	18.6	42.04400	50.0750	
758.1537	18.6	33.00200	39.3059	
401.1434	18.6	18.71700	22.2922	
603.0265	18.6	9.71200	11.5671	
759.1440	18.6	9.39300	11.1872	
195.0628	18.6	4.45000	5.3000	
Comp 94	m=181.0854	Da	t=19.0 min	a=694.9 Kcuentas
181.0854	19.0	694.91400	100.0000	
343.1358	19.0	554.39800	79.7794	
344.1397	18.9	96.11500	13.8312	
182.0885	19.0	77.67000	11.1769	
608.0941	19.0	73.13100	10.5237	
687.2784	19.0	51.86100	7.4629	
457.0771	19.0	43.45200	6.2529	
608.5945	19.0	31.73800	4.5672	
740.1862	19.0	28.44000	4.0926	
541.0208	18.9	27.06200	3.8943	
Comp 95	m=385.1832	Da	t=19.2 min	a=179.0 Kcuentas
385.1832	19.2	179.01900	100.0000	
445.2017	19.2	164.77000	92.0405	
619.1204	19.2	60.23900	33.6495	
785.0805	19.1	53.35800	29.8058	
541.0233	19.2	44.66900	24.9521	
392.0332	19.1	42.36100	23.6629	
446.2059	19.2	37.13400	20.7430	
386.1865	19.2	35.74900	19.9694	
169.0125	19.1	33.90300	18.9382	
499.1219	19.2	24.21400	13.5259	
Comp 97	m=179.0700	Da	t=19.4 min	a=754.9 Kcuentas
179.0700	19.4	754.92600	100.0000	
341.1207	19.4	591.16300	78.3074	
307.1193	19.4	128.08400	16.9664	
342.1236	19.4	103.70400	13.7370	
180.0728	19.4	76.12600	10.0839	
683.2470	19.4	58.40600	7.7367	



455.0600	19.4	54.42500	7.2093		
736.1590	19.4	46.60800	6.1739		
185.0375	19.4	31.71600	4.2012		
541.0213	19.4	29.01900	3.8440		
Comp 100	m=453.1680	Da	t=20.0 min	a=94.7	Kcuentas
453.1680	20.0	94.67400	100.0000		
553.1856	20.1	39.89300	42.1372		
289.0674	20.0	33.26900	35.1406		
599.2274	20.0	28.92800	30.5554		
649.1723	19.9	25.24200	26.6620		
541.0200	20.1	21.76300	22.9873		
454.1714	20.0	17.67800	18.6725		
455.1669	20.0	11.05700	11.6790		
438.0327	20.1	8.43300	8.9074		
600.2285	20.0	5.27200	5.5686		
Comp 102	m=457.0730	Da	t=20.5 min	a=393.0	Kcuentas
457.0730	20.5	393.04900	100.0000		
449.1043	20.5	278.91200	70.9611		
169.0127	20.5	174.30900	44.3479		
458.0770	20.5	101.15200	25.7352		
450.1090	20.5	56.52000	14.3799		
703.1240	20.5	44.00600	11.1961		
759.2255	20.5	43.16300	10.9816		
287.0530	20.5	41.80700	10.6366		
305.0686	20.5	33.47100	8.5157		
553.1842	20.6	24.73700	6.2936		
Comp 105	m=327.1422	Da	t=20.9 min	a=8114.1	Kcuentas
327.1422	20.9	8114.14900	100.0000		
328.1453	20.9	1419.59300	17.4953		
371.0951	20.9	724.27400	8.9261		
655.2890	20.9	586.89200	7.2329		
656.2928	20.9	222.66100	2.7441		
329.1474	20.9	215.60400	2.6571		
708.2007	20.9	184.17500	2.2698		
165.0901	20.9	131.52500	1.6209		
372.0990	20.9	130.70200	1.6108		
709.2030	20.9	63.92200	0.7878		
Comp 109	m=457.1106	Da	t=21.8 min	a=1895.1	Kcuentas
457.1106	21.8	1895.09500	100.0000		
458.1142	21.8	450.73300	23.7842		
459.1135	21.8	163.28500	8.6162		
415.1545	21.7	52.53900	2.7724		
541.0184	21.8	31.40600	1.6572		
460.1143	21.8	30.36200	1.6021		
601.1429	21.7	25.79500	1.3611		
571.0487	21.8	23.13300	1.2207		
Comp 110	m=611.2465	Da	t=22.0 min	a=325.2	Kcuentas
611.2465	22.0	325.17500	100.0000		
616.0152	22.0	169.48400	52.1209		
629.2357	22.0	127.39300	39.1768		
616.5169	22.0	98.40500	30.2622		
612.2513	22.0	97.23100	29.9011		
625.1298	22.0	66.73300	20.5222		
630.2355	22.0	47.10200	14.4851		
617.0149	22.0	45.13900	13.8814		
393.1718	22.0	42.93900	13.2049		
407.1649	22.0	40.85900	12.5652		
Comp 112	m=479.0771	Da	t=22.3 min	a=3237.0	Kcuentas
479.0771	22.3	3236.96600	100.0000		



480.0804	22.3	811.20500	25.0607	
481.0830	22.3	193.21800	5.9691	
959.1620	22.3	63.66000	1.9667	
316.0182	22.3	29.47700	0.9106	
482.0874	22.3	26.64600	0.8232	
960.1645	22.2	25.01200	0.7727	
532.9954	22.3	19.08500	0.5896	
593.0135	22.2	16.82800	0.5199	
478.0707	22.2	15.69400	0.4848	
Comp 114	m=479.0778	Da	t=22.5 min	a=1418.7 Kcuentas
479.0778	22.5	1418.74200	100.0000	
457.1126	22.6	1187.54000	83.7037	
480.0806	22.5	317.74500	22.3962	
458.1171	22.6	272.09300	19.1785	
459.1138	22.6	105.24200	7.4180	
659.2476	22.5	99.74700	7.0307	
481.0843	22.5	69.86900	4.9247	
660.2488	22.5	32.03100	2.2577	
593.0163	22.5	26.05300	1.8363	
541.0198	22.6	24.97400	1.7603	
Comp 115	m=495.1458	Da	t=22.7 min	a=241.7 Kcuentas
495.1458	22.7	241.72700	100.0000	
617.0196	22.8	70.10900	29.0034	
496.1489	22.8	52.05500	21.5346	
537.1894	22.8	22.75400	9.4131	
385.1165	22.7	22.37100	9.2547	
765.2886	22.8	14.98900	6.2008	
617.5205	22.8	13.85500	5.7317	
387.2011	22.8	12.14400	5.0238	
447.2194	22.8	11.48800	4.7525	
Comp 116	m=595.1224	Da	t=22.9 min	a=297.7 Kcuentas
595.1224	22.9	297.69900	100.0000	
616.0162	22.8	266.13100	89.3960	
616.5172	22.8	186.43200	62.6243	
521.1964	23.0	184.36400	61.9297	
433.0384	23.0	140.40600	47.1637	
596.1276	22.9	75.66900	25.4180	
522.1983	23.0	62.36700	20.9497	
519.1939	22.9	62.35600	20.9460	
455.0978	23.0	28.41000	9.5432	
434.0396	22.9	22.80300	7.6598	
Comp 117	m=449.0683	Da	t=23.1 min	a=222.6 Kcuentas
449.0683	23.1	222.63600	100.0000	
465.0987	23.2	58.13900	26.1139	
450.0716	23.1	52.26000	23.4733	
461.1257	23.2	28.08300	12.6139	
455.0966	23.2	17.54300	7.8797	
599.2257	23.1	16.43200	7.3807	
447.2179	23.2	8.72800	3.9203	
169.0133	23.1	8.36400	3.7568	
466.1032	23.2	8.27900	3.7186	
232.9745	23.1	6.55000	2.9420	
Comp 119	m=433.0382	Da	t=23.7 min	a=1062.9 Kcuentas
433.0382	23.7	1062.94700	100.0000	
434.0413	23.7	207.94900	19.5634	
583.1604	23.7	69.59900	6.5477	
521.1941	23.7	51.34000	4.8300	
435.0434	23.6	40.53400	3.8134	
546.9763	23.6	25.61200	2.4095	



522.1990	23.7	19.11500	1.7983		
584.1585	23.7	19.05000	1.7922		
565.2742	23.7	17.65900	1.6613		
605.9878	23.7	10.58300	0.9956		
Comp 121	m=449.0674	Da	t=24.0 min	a=1377.6	Kcuentas
449.0674	24.0	1377.59400	100.0000		
327.1413	23.9	326.77100	23.7204		
450.0707	24.0	302.22500	21.9386		
369.1513	23.9	247.45800	17.9631		
667.2517	24.0	89.91500	6.5270		
309.1303	23.9	85.05900	6.1745		
581.2154	24.0	82.30700	5.9747		
451.0719	24.0	67.28600	4.8843		
328.1458	23.9	54.95300	3.9891		
370.1552	23.9	42.24400	3.0665		
Comp 126	m=463.0828	Da	t=24.7 min	a=2774.5	Kcuentas
463.0828	24.7	2774.50800	100.0000		
597.1745	24.7	1049.34100	37.8208		
464.0863	24.7	670.96700	24.1833		
300.9965	24.7	469.12800	16.9085		
598.1782	24.7	321.68800	11.5944		
465.0889	24.7	138.46900	4.9908		
927.1725	24.7	94.63400	3.4108		
451.0965	24.7	79.82100	2.8769		
599.1802	24.7	79.06100	2.8496		
301.9995	24.7	71.38600	2.5729		
Comp 128	m=463.0827	Da	t=25.0 min	a=1068.9	Kcuentas
463.0827	25.0	1068.88400	100.0000		
464.0863	25.0	237.66200	22.2346		
599.0964	25.0	118.01900	11.0413		
305.0657	25.0	97.47000	9.1189		
465.0877	25.0	56.02800	5.2417		
441.0865	25.0	48.07800	4.4980		
600.1005	25.0	36.44300	3.4094		
541.0178	25.0	21.77200	2.0369		
625.1169	25.0	15.59100	1.4586		
967.3329	25.1	11.83800	1.1075		
Comp 130	m=523.2104	Da	t=25.6 min	a=168.5	Kcuentas
523.2104	25.6	168.47900	100.0000		
625.1122	25.7	57.76200	34.2844		
524.2144	25.6	56.20200	33.3585		
625.6119	25.6	28.66300	17.0128		
303.0460	25.7	22.77500	13.5180		
429.2066	25.7	13.64200	8.0972		
525.2143	25.6	11.93100	7.0816		
626.1108	25.6	11.82300	7.0175		
521.1911	25.6	6.47000	3.8402		
Comp 131	m=433.0718	Da	t=25.9 min	a=371.9	Kcuentas
433.0718	25.9	371.93100	100.0000		
523.2102	25.8	164.96600	44.3539		
434.0749	25.9	84.26700	22.6566		
524.2137	25.8	42.64600	11.4661		
491.1841	25.8	28.48500	7.6587		
573.2259	25.9	26.35000	7.0846		
547.0071	25.9	13.33000	3.5840		
435.0788	25.9	12.80900	3.4439		
300.0232	25.9	11.97000	3.2183		
597.1754	25.9	11.86100	3.1890		
Comp 134	m=433.0728	Da	t=26.4 min	a=1474.0	Kcuentas



433.0728	26.4	1473.95400	100.0000	
447.0881	26.3	445.25000	30.2079	
434.0760	26.4	330.33200	22.4113	
448.0928	26.3	112.26300	7.6165	
435.0855	26.4	82.68600	5.6098	
300.0239	26.4	62.30800	4.2273	
867.1538	26.3	34.74300	2.3571	
301.0250	26.3	22.88300	1.5525	
449.0904	26.3	21.66700	1.4700	
547.0124	26.4	20.56600	1.3953	
Comp 136	m=479.1492	Da	t=26.6 min	a=148.4 Kcuentas
479.1492	26.6	148.44100	100.0000	
480.1518	26.6	38.70800	26.0764	
232.9737	26.6	21.41600	14.4273	
Comp 138	m=433.0710	Da	t=27.0 min	a=2185.7 Kcuentas
433.0710	27.0	2185.72800	100.0000	
434.0740	27.0	509.34700	23.3033	
300.0224	27.0	98.34800	4.4996	
867.1496	26.9	96.44400	4.4124	
435.0777	27.0	94.69600	4.3325	
617.1153	27.0	94.59700	4.3279	
301.0285	27.0	75.83100	3.4694	
617.6139	26.9	51.05100	2.3357	
547.0111	27.0	50.41700	2.3066	
807.2984	26.9	41.70500	1.9081	
Comp 140	m=565.2195	Da	t=27.3 min	a=350.2 Kcuentas
565.2195	27.3	350.17300	100.0000	
439.1008	27.4	349.94400	99.9346	
609.1164	27.3	118.45400	33.8273	
566.2225	27.3	97.14600	27.7423	
807.2980	27.3	73.11400	20.8794	
448.0900	27.3	42.06500	12.0126	
747.2764	27.4	34.87500	9.9594	
585.2232	27.3	32.47000	9.2726	
808.2960	27.3	32.42100	9.2586	
603.2322	27.3	24.81100	7.0854	
Comp 144	m=537.1531	Da	t=27.9 min	a=119.1 Kcuentas
537.1531	27.9	119.05700	100.0000	
427.1544	27.8	97.15400	81.6029	
538.1553	27.8	34.47100	28.9534	
428.1572	27.8	17.55700	14.7467	
430.1692	27.8	16.32300	13.7102	
249.1065	27.8	11.78500	9.8986	
405.1142	27.9	8.39200	7.0487	
Comp 149	m=289.1071	Da	t=28.7 min	a=72.2 Kcuentas
289.1071	28.7	72.21100	100.0000	
747.2739	28.8	27.93900	38.6908	
232.9702	28.7	20.63200	28.5718	
431.1867	28.8	19.50500	27.0111	
442.1196	28.8	8.96900	12.4205	
Comp 153	m=435.1121	Da	t=29.4 min	a=288.2 Kcuentas
435.1121	29.4	288.23100	100.0000	
436.1208	29.4	58.80800	20.4031	
505.1992	29.4	48.36900	16.7813	
273.0713	29.3	39.27000	13.6245	
232.9716	29.3	38.01900	13.1905	
417.0745	29.3	21.62800	7.5037	
431.1848	29.4	18.77900	6.5153	
437.1249	29.3	10.00200	3.4701	



Comp 156	m=405.1130	Da	t=29.9 min	a=890.0	Kcuentas
405.1130	29.9	890.02200	100.0000		
406.1163	29.9	181.46200	20.3885		
625.1090	29.9	177.21100	19.9109		
435.0873	29.8	154.14300	17.3190		
297.0578	29.9	81.45500	9.1520		
626.1105	29.9	54.14500	6.0836		
399.1235	29.8	39.77700	4.4692		
407.1206	29.9	37.48400	4.2116		
436.0913	29.8	32.75500	3.6802		
745.2586	29.8	24.84000	2.7909		
Comp 158	m=419.0939	Da	t=31.0 min	a=538.6	Kcuentas
419.0939	31.0	538.59200	100.0000		
420.0967	31.0	116.82200	21.6903		
447.2168	30.9	43.17800	8.0168		
625.1077	31.0	28.00900	5.2004		
533.0333	30.9	25.40200	4.7164		
232.9716	31.0	23.16700	4.3014		
421.1007	31.0	16.86600	3.1315		
626.1113	30.9	8.81100	1.6359		
Comp 159	m=609.1162	Da	t=31.6 min	a=129.6	Kcuentas
609.1162	31.6	129.64600	100.0000		
232.9713	31.6	48.78900	37.6325		
610.1216	31.6	45.00500	34.7138		
553.1484	31.7	12.82000	9.8885		
611.1216	31.6	5.38000	4.1498		
Comp 161	m=609.1150	Da	t=32.0 min	a=456.0	Kcuentas
609.1150	32.0	456.00500	100.0000		
610.1189	32.0	139.31700	30.5516		
611.1189	32.0	30.26300	6.6366		
615.2321	32.1	25.68600	5.6328		
Comp 164	m=429.0979	Da	t=32.8 min	a=76.2	Kcuentas
429.0979	32.8	76.22200	100.0000		
537.1567	32.8	59.66100	78.2727		
232.9718	32.8	19.22200	25.2184		
538.1564	32.8	14.25600	18.7033		
430.1030	32.8	11.94200	15.6674		
Comp 165	m=232.9723	Da	t=33.2 min	a=112.8	Kcuentas
232.9723	33.2	112.84200	100.0000		
609.1130	33.2	68.24900	60.4819		
Comp 169	m=593.1189	Da	t=34.0 min	a=893.7	Kcuentas
593.1189	34.0	893.68200	100.0000		
594.1212	34.0	272.24900	30.4637		
595.1244	34.0	56.56400	6.3293		
311.1616	34.0	36.89400	4.1283		
707.0585	34.0	22.05700	2.4681		
537.1543	34.0	17.28000	1.9336		
Comp 171	m=593.1186	Da	t=34.7 min	a=332.5	Kcuentas
593.1186	34.7	332.54000	100.0000		
594.1230	34.7	92.50200	27.8168		
595.1259	34.7	20.68200	6.2194		
Comp 178	m=327.2117	Da	t=36.7 min	a=115.0	Kcuentas
327.2117	36.7	114.96600	100.0000		
232.9724	36.7	91.33000	79.4409		
325.1782	36.8	17.95700	15.6194		
312.1661	36.7	10.74900	9.3497		



APÉNDICE 3: ACEITE.

Se ha elegido una muestra representante del grupo de aceite con las siguientes características:

Nombre de la muestra	gg1.bin
Intervalo masas analizadas	[0 Da, 900 Da]
Intervalo de tiempo analizado	[0 min, 30 min]
Tamaño fichero	2 MB

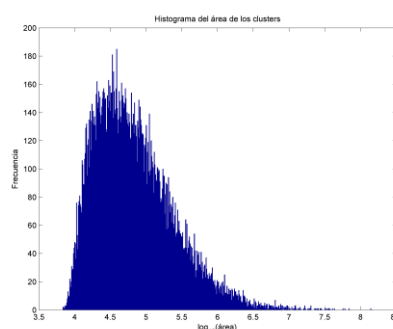
Se ha realizado la transformación del formato baf a XML, después se ha convertido en formato binario, se ha eliminado el ruido y se ha calibrado en masa. Se ha realizado el *clustering* de líneas utilizando los siguientes parámetros:

Nombre del parámetro	Valor utilizado
Umbral intensidad	100 cuentas
Umbral área	100
Resolución en tiempo fase 1	300s
Resolución en tiempo fase 2	20s
Umbral tiempo de retención	20s
Umbral m/z	0.010Da
Umbral área huella	10 ⁴
Línea de fondo	si



Número de *clusters*, tiempo computacional e histograma del área de los *clusters* obtenidos con el *clustering* de líneas para la muestra representativa de aceite:

Clustering de líneas	Nº de clusters
Número inicial de puntos (t,m/z,l) sin ruido	1.532.149
Número de puntos (t,m/z,l) sin ruido, umbral intensidad =100	278.483
<i>Pre-Clustering</i> fase1: resolución tiempo =300s	5.722
<i>Pre-Clustering</i> fase2: : resolución tiempo =20s	13.347
Elimina: puntos<3, area<100	4.148
<i>Clustering</i> de puntos	22.538
Elimina: puntos<3, area<100	6.584
Detección de picos cromatográficos	26.524
Elimina: puntos<3, area<100	22.913
Umbral de area> 1000	22.625
Umbral de area> 10 ⁴	7.673
Umbral de area> 10 ⁵	835
Umbral de area> 10 ⁶	51



Tiempo computacional

<i>Preclustering</i> : fase 1 y 2	1.9 min	24.4%
<i>Clustering</i> de puntos	2.0 min	25.6%
Detección de picos cromatográficos	2.7 min	34.6%
Estadística	1.2 min	15.4%
TOTAL	7.8 min	100%



Compuestos identificados por el método convencional en los extractos polifenólico de aceite de oliva:

RT	m/z	[M-H] ⁻	Compuestos
6,47	153,0555	C ₈ H ₉ O ₃	Hidroxitirosol
8,15	137,0605	C ₈ H ₉ O ₂	Tirosol
14,53	319,1177	C ₁₉ H ₁₇ O ₆	Deacetoxi oleuropeína aglicona
15,94	285,0399	C ₁₅ H ₉ O ₆	Luteolina
16,69	357,1339	C ₂₀ H ₂₁ O ₆	Pinoresinol
17,33	415,1381	C ₂₂ H ₂₃ O ₈	Acetoxipinoresinol
18,68	269,0468	C ₁₅ H ₉ O ₅	Apigenin
17,71	303,1237	C ₁₇ H ₁₉ O ₅	Deacetoxi ligustrósido aglicona
14,12	377,1231	C ₁₉ H ₂₁ O ₈	Oleuropeína aglicona isómero 1
14,78	377,1246	C ₁₉ H ₂₁ O ₉	Oleuropeína aglicona isómero 2
16,22	377,1236	C ₁₉ H ₂₁ O ₁₀	Oleuropeína aglicona isómero 3
17,41	377,1260	C ₁₉ H ₂₁ O ₁₁	Oleuropeína aglicona isómero 4
17,88	377,1253	C ₁₉ H ₂₁ O ₁₂	Oleuropeína aglicona isómero 5
18,88	377,1264	C ₁₉ H ₂₁ O ₁₃	Oleuropeína aglicona isómero 6
19,05	377,1268	C ₁₉ H ₂₁ O ₁₄	Oleuropeína aglicona isómero 7
20,89	377,1271	C ₁₉ H ₂₁ O ₁₅	Oleuropeína aglicona isómero 8
21,02	377,1267	C ₁₉ H ₂₁ O ₁₆	Oleuropeína aglicona isómero 9
21,59	377,1269	C ₁₉ H ₂₁ O ₁₇	Oleuropeína aglicona isómero 10
21,87	377,1259	C ₁₉ H ₂₁ O ₁₈	Oleuropeína aglicona isómero 11
7,83	389,1076	C ₁₆ H ₂₁ O ₁₁	Secologanósido u oleósido
9,00	389,1456	C ₁₇ H ₂₅ O ₁₀	Secologanol
13,15	187,0975	C ₉ H ₁₅ O ₄	Ácido azeláico
13,88	335,1145	C ₁₇ H ₁₉ O ₇	Deacetoxi 10-hidroxi-oleuropeína aglicona
15,72	417,1530	C ₂₂ H ₂₅ O ₈	Siringaresinol
16,99	391,1387	C ₂₀ H ₂₃ O ₈	Metil oleuropeína aglicona isómero 1
19,30	299,0577	C ₁₆ H ₁₁ O ₆	Crisoeriol
23,03	391,1412	C ₂₀ H ₂₃ O ₈	Metil oleuropeína aglicona isómero 2
24,39	227,2016	C ₁₄ H ₂₇ O ₂	Ácido mirístico
16,47	361,1286	C ₁₉ H ₂₁ O ₇	Ligustrósido aglicona isómero 1
16,64	361,1284	C ₁₉ H ₂₁ O ₈	Ligustrósido aglicona isómero 2
17,48	361,1299	C ₁₉ H ₂₁ O ₉	Ligustrósido aglicona isómero 3
19,10	361,1312	C ₁₉ H ₂₁ O ₁₀	Ligustrósido aglicona isómero 4
20,74	361,1315	C ₁₉ H ₂₁ O ₁₁	Ligustrósido aglicona isómero 5
20,99	361,1360	C ₁₉ H ₂₁ O ₁₂	Ligustrósido aglicona isómero 6
21,89	361,1305	C ₁₉ H ₂₁ O ₁₃	Ligustrósido aglicona isómero 7
22,04	361,1305	C ₁₉ H ₂₁ O ₁₄	Ligustrósido aglicona isómero 8
23,33	361,1300	C ₁₉ H ₂₁ O ₁₅	Ligustrósido aglicona isómero 9
23,58	361,1289	C ₁₉ H ₂₁ O ₁₆	Ligustrósido aglicona isómero 10



Lista de compuestos obtenidos con el *clustering* de líneas para la muestra representativa de aceite:

Muestra: 1. Fichero con clusters: cluster_estad_aceite_gg1_eliminal000.m
 Umbral area: 200000 cuentas

Eliminados 102 clusters alrededor de la masa 253.2154 Da: t_tot=383.6 s de 1495.6 s
 Eliminados 122 clusters alrededor de la masa 141.0169 Da: t_tot=394.7 s de 1495.6 s
 Eliminados 123 clusters alrededor de la masa 281.2471 Da: t_tot=413.8 s de 1495.6 s
 Eliminados 118 clusters alrededor de la masa 255.2305 Da: t_tot=413.8 s de 1495.6 s
 Eliminados 116 clusters alrededor de la masa 96.9572 Da: t_tot=418.0 s de 1495.6 s
 Eliminados 101 clusters alrededor de la masa 217.0023 Da: t_tot=401.1 s de 1495.6 s
 Eliminados 113 clusters alrededor de la masa 125.9953 Da: t_tot=394.5 s de 1495.6 s
 Eliminados 129 clusters alrededor de la masa 279.2296 Da: t_tot=409.2 s de 1495.6 s
 Eliminados 119 clusters alrededor de la masa 89.0204 Da: t_tot=405.9 s de 1495.6 s
 Eliminados 102 clusters alrededor de la masa 174.9925 Da: t_tot=380.2 s de 1495.6 s
 Eliminados 124 clusters alrededor de la masa 59.0096 Da: t_tot=417.3 s de 1495.6 s
 Eliminados 115 clusters alrededor de la masa 201.0261 Da: t_tot=401.2 s de 1495.6 s
 Eliminados 111 clusters alrededor de la masa 291.9881 Da: t_tot=378.5 s de 1495.6 s
 Eliminados 118 clusters alrededor de la masa 283.2614 Da: t_tot=408.0 s de 1495.6 s
 Eliminados 122 clusters alrededor de la masa 203.0172 Da: t_tot=403.0 s de 1495.6 s
 Eliminados 118 clusters alrededor de la masa 128.0374 Da: t_tot=400.1 s de 1495.6 s
 Eliminados 113 clusters alrededor de la masa 247.0128 Da: t_tot=376.8 s de 1495.6 s
 Eliminados 115 clusters alrededor de la masa 173.0128 Da: t_tot=392.0 s de 1495.6 s
 Eliminados 124 clusters alrededor de la masa 232.0411 Da: t_tot=393.3 s de 1495.6 s

Numero de clusters iniciales: 12812

Lista de compuestos ordenados por tiempo de retención

Comp.1	m= 112.9852	Da	se(m)= 0.2	mDa	t= 0.40 min	t= 24.2 s	se(t)= 2.5 s	a= 1698.2	Kcuentas tiene 300 líneas
Comp.2	m= 127.0010	Da	se(m)= 0.8	mDa	t= 0.50 min	t= 30.2 s	se(t)= 5.3 s	a= 323.4	Kcuentas tiene 26 líneas
Comp.3	m= 387.0262	Da	se(m)= 1.3	mDa	t= 0.69 min	t= 41.3 s	se(t)=10.1 s	a= 251.6	Kcuentas tiene 34 líneas
Comp.4	m= 123.0444	Da	se(m)= 0.4	mDa	t= 6.47 min	t=388.0 s	se(t)= 5.2 s	a= 1241.0	Kcuentas tiene 40 líneas
Comp.5	m= 137.0595	Da	se(m)= 1.1	mDa	t= 8.18 min	t=491.0 s	se(t)= 4.3 s	a= 210.7	Kcuentas tiene 31 líneas
Comp.6	m= 151.0396	Da	se(m)= 0.8	mDa	t=10.01 min	t=600.5 s	se(t)=12.7 s	a= 362.4	Kcuentas tiene 5 líneas
Comp.7	m= 151.0392	Da	se(m)= 0.6	mDa	t=10.41 min	t=624.5 s	se(t)= 4.3 s	a= 257.4	Kcuentas tiene 41 líneas
Comp.8	m= 151.0395	Da	se(m)= 0.5	mDa	t=10.65 min	t=639.2 s	se(t)= 4.2 s	a= 257.8	Kcuentas tiene 32 líneas
Comp.9	m= 151.0391	Da	se(m)= 1.0	mDa	t=11.02 min	t=661.3 s	se(t)=11.8 s	a= 324.4	Kcuentas tiene 10 líneas
Comp.10	m= 139.0048	Da	se(m)= 1.3	mDa	t=12.30 min	t=738.1 s	se(t)= 9.3 s	a= 210.7	Kcuentas tiene 19 líneas



Comp.11	m= 349.1287	Da	se(m)= 0.8 mDa	t=12.62 min	t=757.1 s	se(t)= 5.1 s	a= 379.0	Kcuentas tiene	52 líneas
Comp.12	m= 377.1234	Da	se(m)= 1.1 mDa	t=12.71 min	t=762.3 s	se(t)= 7.7 s	a= 605.6	Kcuentas tiene	44 líneas
Comp.13	m= 349.1289	Da	se(m)= 0.7 mDa	t=12.83 min	t=769.7 s	se(t)= 2.5 s	a= 221.2	Kcuentas tiene	85 líneas
Comp.14	m= 349.1292	Da	se(m)= 0.7 mDa	t=13.12 min	t=787.3 s	se(t)= 7.7 s	a= 616.1	Kcuentas tiene	46 líneas
Comp.15	m= 377.1237	Da	se(m)= 0.6 mDa	t=13.26 min	t=795.9 s	se(t)= 8.6 s	a= 908.9	Kcuentas tiene	35 líneas
Comp.16	m= 275.0897	Da	se(m)= 0.7 mDa	t=13.35 min	t=801.0 s	se(t)= 4.1 s	a= 215.0	Kcuentas tiene	68 líneas
Comp.17	m= 139.0066	Da	se(m)= 0.8 mDa	t=13.40 min	t=804.0 s	se(t)=11.2 s	a= 1383.8	Kcuentas tiene	28 líneas
Comp.18	m= 127.0396	Da	se(m)= 0.6 mDa	t=13.51 min	t=810.8 s	se(t)= 8.0 s	a= 461.8	Kcuentas tiene	72 líneas
Comp.19	m= 377.1238	Da	se(m)= 0.4 mDa	t=13.61 min	t=816.5 s	se(t)= 3.4 s	a= 384.7	Kcuentas tiene	73 líneas
Comp.20	m= 301.1074	Da	se(m)= 1.5 mDa	t=13.62 min	t=817.3 s	se(t)=16.1 s	a= 259.2	Kcuentas tiene	2 líneas
Comp.21	m= 319.1184	Da	se(m)= 0.4 mDa	t=13.71 min	t=822.6 s	se(t)=18.9 s	a= 8948.6	Kcuentas tiene	7 líneas
Comp.22	m= 265.1472	Da	se(m)= 0.7 mDa	t=13.74 min	t=824.5 s	se(t)= 5.9 s	a= 253.0	Kcuentas tiene	57 líneas
Comp.23	m= 639.2449	Da	se(m)= 1.0 mDa	t=13.81 min	t=828.4 s	se(t)=15.2 s	a= 2998.6	Kcuentas tiene	11 líneas
Comp.24	m= 335.1136	Da	se(m)= 1.0 mDa	t=13.89 min	t=833.1 s	se(t)= 4.6 s	a= 676.7	Kcuentas tiene	98 líneas
Comp.25	m= 95.0466	Da	se(m)= 0.3 mDa	t=14.09 min	t=845.2 s	se(t)= 5.0 s	a= 532.7	Kcuentas tiene	84 líneas
Comp.26	m= 377.1231	Da	se(m)= 0.6 mDa	t=14.19 min	t=851.6 s	se(t)= 7.0 s	a= 1948.1	Kcuentas tiene	79 líneas
Comp.27	m= 183.0656	Da	se(m)= 0.2 mDa	t=14.35 min	t=860.9 s	se(t)= 7.6 s	a= 4599.6	Kcuentas tiene	85 líneas
Comp.28	m= 95.0465	Da	se(m)= 0.2 mDa	t=14.45 min	t=867.1 s	se(t)= 4.3 s	a= 493.5	Kcuentas tiene	67 líneas
Comp.29	m= 319.1184	Da	se(m)= 0.4 mDa	t=14.46 min	t=867.8 s	se(t)=10.8 s	a= 12193.0	Kcuentas tiene	28 líneas
Comp.30	m= 183.0660	Da	se(m)= 0.2 mDa	t=14.68 min	t=880.6 s	se(t)= 4.1 s	a= 2542.7	Kcuentas tiene	100 líneas
Comp.31	m= 377.1247	Da	se(m)= 1.0 mDa	t=14.85 min	t=890.7 s	se(t)= 8.0 s	a= 4665.2	Kcuentas tiene	100 líneas
Comp.32	m= 640.2482	Da	se(m)= 0.5 mDa	t=14.94 min	t=896.5 s	se(t)= 4.2 s	a= 847.5	Kcuentas tiene	73 líneas
Comp.33	m= 301.1075	Da	se(m)= 0.9 mDa	t=15.06 min	t=903.5 s	se(t)= 9.2 s	a= 311.4	Kcuentas tiene	30 líneas
Comp.34	m= 265.1474	Da	se(m)= 0.6 mDa	t=15.19 min	t=911.4 s	se(t)= 4.3 s	a= 289.5	Kcuentas tiene	74 líneas
Comp.35	m= 319.1189	Da	se(m)= 0.4 mDa	t=15.21 min	t=912.8 s	se(t)=17.2 s	a= 13962.3	Kcuentas tiene	9 líneas
Comp.36	m= 275.0907	Da	se(m)= 0.7 mDa	t=15.30 min	t=918.1 s	se(t)= 4.3 s	a= 310.9	Kcuentas tiene	107 líneas
Comp.37	m= 320.1222	Da	se(m)= 0.6 mDa	t=15.33 min	t=919.8 s	se(t)=14.5 s	a= 1824.0	Kcuentas tiene	17 líneas
Comp.38	m= 377.1237	Da	se(m)= 0.7 mDa	t=15.51 min	t=930.5 s	se(t)= 7.0 s	a= 637.1	Kcuentas tiene	93 líneas
Comp.39	m= 275.0901	Da	se(m)= 1.0 mDa	t=15.62 min	t=937.3 s	se(t)= 6.9 s	a= 416.2	Kcuentas tiene	55 líneas
Comp.40	m= 377.1245	Da	se(m)= 0.7 mDa	t=15.84 min	t=950.3 s	se(t)= 4.3 s	a= 375.0	Kcuentas tiene	87 líneas
Comp.41	m= 285.0401	Da	se(m)= 0.6 mDa	t=15.97 min	t=958.1 s	se(t)= 3.6 s	a= 1296.0	Kcuentas tiene	100 líneas
Comp.42	m= 391.1402	Da	se(m)= 0.9 mDa	t=16.06 min	t=963.3 s	se(t)=11.2 s	a= 316.2	Kcuentas tiene	1 líneas
Comp.43	m= 265.1470	Da	se(m)= 0.7 mDa	t=16.09 min	t=965.6 s	se(t)= 4.3 s	a= 391.2	Kcuentas tiene	66 líneas
Comp.44	m= 377.1249	Da	se(m)= 0.7 mDa	t=16.26 min	t=975.4 s	se(t)= 6.6 s	a= 2260.4	Kcuentas tiene	92 líneas
Comp.45	m= 265.1472	Da	se(m)= 0.6 mDa	t=16.37 min	t=982.2 s	se(t)= 5.2 s	a= 494.2	Kcuentas tiene	67 líneas
Comp.46	m= 183.0661	Da	se(m)= 0.6 mDa	t=16.46 min	t=987.6 s	se(t)= 7.7 s	a= 456.0	Kcuentas tiene	20 líneas
Comp.47	m= 361.1294	Da	se(m)= 0.8 mDa	t=16.58 min	t=994.7 s	se(t)= 9.4 s	a= 1250.6	Kcuentas tiene	49 líneas
Comp.48	m= 265.1473	Da	se(m)= 0.6 mDa	t=16.67 min	t=1000.1 s	se(t)= 5.1 s	a= 515.8	Kcuentas tiene	105 líneas
Comp.49	m= 183.0660	Da	se(m)= 0.5 mDa	t=16.91 min	t=1014.5 s	se(t)= 7.7 s	a= 673.3	Kcuentas tiene	73 líneas



Comp.50	m= 303.1225	Da	se(m)= 0.9 mDa	t=17.08 min	t=1024.8 s	se(t)=13.2 s	a= 441.4	Kcuentas tiene	1 líneas
Comp.51	m= 377.1234	Da	se(m)= 0.5 mDa	t=17.09 min	t=1025.5 s	se(t)= 4.0 s	a= 518.2	Kcuentas tiene	134 líneas
Comp.52	m= 183.0659	Da	se(m)= 0.4 mDa	t=17.32 min	t=1039.4 s	se(t)= 4.2 s	a= 420.9	Kcuentas tiene	96 líneas
Comp.53	m= 345.0970	Da	se(m)= 0.9 mDa	t=17.41 min	t=1044.6 s	se(t)= 9.2 s	a= 342.4	Kcuentas tiene	11 líneas
Comp.54	m= 265.1468	Da	se(m)= 0.5 mDa	t=17.45 min	t=1046.7 s	se(t)= 2.6 s	a= 304.9	Kcuentas tiene	31 líneas
Comp.55	m= 361.1290	Da	se(m)= 0.6 mDa	t=17.51 min	t=1050.6 s	se(t)= 7.5 s	a= 2092.2	Kcuentas tiene	109 líneas
Comp.56	m= 183.0653	Da	se(m)= 0.5 mDa	t=17.68 min	t=1060.9 s	se(t)= 7.3 s	a= 801.8	Kcuentas tiene	76 líneas
Comp.57	m= 265.1467	Da	se(m)= 0.4 mDa	t=17.84 min	t=1070.6 s	se(t)= 2.5 s	a= 299.0	Kcuentas tiene	50 líneas
Comp.58	m= 377.1242	Da	se(m)= 0.7 mDa	t=17.93 min	t=1075.7 s	se(t)= 6.7 s	a= 3825.1	Kcuentas tiene	118 líneas
Comp.59	m= 265.1466	Da	se(m)= 0.4 mDa	t=18.05 min	t=1083.0 s	se(t)= 4.3 s	a= 541.0	Kcuentas tiene	78 líneas
Comp.60	m= 183.0658	Da	se(m)= 0.3 mDa	t=18.30 min	t=1097.8 s	se(t)= 6.8 s	a= 699.2	Kcuentas tiene	86 líneas
Comp.61	m= 265.1468	Da	se(m)= 0.5 mDa	t=18.42 min	t=1105.4 s	se(t)= 1.7 s	a= 235.2	Kcuentas tiene	30 líneas
Comp.62	m= 377.1242	Da	se(m)= 0.6 mDa	t=18.47 min	t=1108.1 s	se(t)= 8.5 s	a= 2245.2	Kcuentas tiene	66 líneas
Comp.63	m= 307.0815	Da	se(m)= 0.5 mDa	t=18.52 min	t=1111.1 s	se(t)= 2.5 s	a= 285.0	Kcuentas tiene	82 líneas
Comp.64	m= 165.0550	Da	se(m)= 0.7 mDa	t=18.58 min	t=1114.7 s	se(t)= 7.5 s	a= 212.3	Kcuentas tiene	5 líneas
Comp.65	m= 265.1469	Da	se(m)= 0.7 mDa	t=18.69 min	t=1121.4 s	se(t)= 7.4 s	a= 1036.9	Kcuentas tiene	67 líneas
Comp.66	m= 311.1680	Da	se(m)= 0.9 mDa	t=18.77 min	t=1126.5 s	se(t)= 5.1 s	a= 233.9	Kcuentas tiene	56 líneas
Comp.67	m= 377.1245	Da	se(m)= 0.5 mDa	t=18.95 min	t=1136.9 s	se(t)= 7.1 s	a= 6055.1	Kcuentas tiene	115 líneas
Comp.68	m= 265.1471	Da	se(m)= 0.5 mDa	t=19.17 min	t=1149.9 s	se(t)= 8.6 s	a= 1256.2	Kcuentas tiene	54 líneas
Comp.69	m= 275.0889	Da	se(m)= 0.4 mDa	t=19.30 min	t=1158.1 s	se(t)= 2.5 s	a= 375.5	Kcuentas tiene	108 líneas
Comp.70	m= 291.0869	Da	se(m)= 0.8 mDa	t=19.39 min	t=1163.3 s	se(t)= 5.1 s	a= 232.7	Kcuentas tiene	16 líneas
Comp.71	m= 307.0814	Da	se(m)= 0.3 mDa	t=19.48 min	t=1168.5 s	se(t)= 6.0 s	a= 1177.6	Kcuentas tiene	103 líneas
Comp.72	m= 308.0851	Da	se(m)= 0.8 mDa	t=19.58 min	t=1174.8 s	se(t)= 7.7 s	a= 274.9	Kcuentas tiene	25 líneas
Comp.73	m= 378.1277	Da	se(m)= 0.6 mDa	t=19.60 min	t=1176.2 s	se(t)= 2.6 s	a= 252.2	Kcuentas tiene	44 líneas
Comp.74	m= 377.1246	Da	se(m)= 0.4 mDa	t=19.67 min	t=1180.0 s	se(t)=14.8 s	a= 7197.3	Kcuentas tiene	10 líneas
Comp.75	m= 307.0817	Da	se(m)= 0.3 mDa	t=19.72 min	t=1183.3 s	se(t)= 2.6 s	a= 565.7	Kcuentas tiene	132 líneas
Comp.76	m= 139.0359	Da	se(m)= 0.6 mDa	t=19.83 min	t=1189.7 s	se(t)= 6.0 s	a= 409.2	Kcuentas tiene	12 líneas
Comp.77	m= 265.1474	Da	se(m)= 0.5 mDa	t=19.91 min	t=1194.8 s	se(t)= 3.5 s	a= 533.3	Kcuentas tiene	128 líneas
Comp.78	m= 149.0239	Da	se(m)= 0.4 mDa	t=20.02 min	t=1201.4 s	se(t)= 4.3 s	a= 393.4	Kcuentas tiene	54 líneas
Comp.79	m= 275.0899	Da	se(m)= 0.5 mDa	t=20.14 min	t=1208.5 s	se(t)= 6.9 s	a= 1548.3	Kcuentas tiene	90 líneas
Comp.80	m= 149.0240	Da	se(m)= 0.4 mDa	t=20.25 min	t=1215.0 s	se(t)= 3.4 s	a= 382.7	Kcuentas tiene	80 líneas
Comp.81	m= 265.1477	Da	se(m)= 0.4 mDa	t=20.39 min	t=1223.3 s	se(t)= 2.6 s	a= 430.7	Kcuentas tiene	112 líneas
Comp.82	m= 101.0220	Da	se(m)= 0.5 mDa	t=20.55 min	t=1232.8 s	se(t)= 9.6 s	a= 294.6	Kcuentas tiene	30 líneas
Comp.83	m= 361.1299	Da	se(m)= 0.7 mDa	t=20.64 min	t=1238.2 s	se(t)= 8.9 s	a= 1399.9	Kcuentas tiene	50 líneas
Comp.84	m= 307.0820	Da	se(m)= 0.4 mDa	t=20.66 min	t=1239.5 s	se(t)=22.5 s	a= 10599.2	Kcuentas tiene	7 líneas
Comp.85	m= 377.1263	Da	se(m)= 1.3 mDa	t=20.75 min	t=1245.2 s	se(t)=17.2 s	a= 26611.4	Kcuentas tiene	16 líneas
Comp.86	m= 116.9278	Da	se(m)= 0.5 mDa	t=20.78 min	t=1247.1 s	se(t)= 5.9 s	a= 295.7	Kcuentas tiene	55 líneas
Comp.87	m= 753.2402	Da	se(m)= 1.3 mDa	t=20.88 min	t=1252.6 s	se(t)= 4.0 s	a= 901.7	Kcuentas tiene	71 líneas
Comp.88	m= 345.0979	Da	se(m)= 0.6 mDa	t=20.89 min	t=1253.6 s	se(t)=10.1 s	a= 2270.9	Kcuentas tiene	83 líneas



Comp.89	m= 361.1294	Da	se(m)= 0.8 mDa	t=21.02 min	t=1261.4 s	se(t)= 3.8 s	a= 234.7	Kcuentas tiene	91 líneas
Comp.90	m= 265.1469	Da	se(m)= 0.4 mDa	t=21.16 min	t=1269.5 s	se(t)= 9.5 s	a= 2086.9	Kcuentas tiene	54 líneas
Comp.91	m= 311.1678	Da	se(m)= 0.6 mDa	t=21.25 min	t=1275.3 s	se(t)=11.7 s	a= 1260.9	Kcuentas tiene	15 líneas
Comp.92	m= 291.0862	Da	se(m)= 0.9 mDa	t=21.36 min	t=1281.3 s	se(t)= 9.1 s	a= 376.7	Kcuentas tiene	30 líneas
Comp.93	m= 377.1242	Da	se(m)= 0.6 mDa	t=21.58 min	t=1294.7 s	se(t)= 5.3 s	a= 2112.5	Kcuentas tiene	143 líneas
Comp.94	m= 265.1469	Da	se(m)= 0.4 mDa	t=21.73 min	t=1303.7 s	se(t)=10.4 s	a= 2828.4	Kcuentas tiene	42 líneas
Comp.95	m= 339.1992	Da	se(m)= 1.0 mDa	t=21.85 min	t=1311.2 s	se(t)= 7.7 s	a= 649.0	Kcuentas tiene	21 líneas
Comp.96	m= 377.1240	Da	se(m)= 0.6 mDa	t=21.94 min	t=1316.3 s	se(t)= 7.4 s	a= 954.6	Kcuentas tiene	80 líneas
Comp.97	m= 433.2354	Da	se(m)= 0.6 mDa	t=22.07 min	t=1324.0 s	se(t)=14.5 s	a= 2186.1	Kcuentas tiene	12 líneas
Comp.98	m= 265.1468	Da	se(m)= 0.4 mDa	t=22.09 min	t=1325.7 s	se(t)= 2.6 s	a= 783.0	Kcuentas tiene	124 líneas
Comp.99	m= 265.1468	Da	se(m)= 0.4 mDa	t=22.27 min	t=1336.2 s	se(t)= 3.5 s	a= 1059.4	Kcuentas tiene	131 líneas
Comp.100	m= 339.1994	Da	se(m)= 0.9 mDa	t=22.37 min	t=1342.3 s	se(t)= 3.4 s	a= 382.5	Kcuentas tiene	31 líneas
Comp.101	m= 265.1472	Da	se(m)= 0.5 mDa	t=22.47 min	t=1348.3 s	se(t)= 3.4 s	a= 1061.4	Kcuentas tiene	150 líneas
Comp.102	m= 571.2887	Da	se(m)= 1.0 mDa	t=22.56 min	t=1353.4 s	se(t)=10.4 s	a= 852.1	Kcuentas tiene	16 líneas
Comp.103	m= 265.1472	Da	se(m)= 0.5 mDa	t=22.62 min	t=1357.2 s	se(t)= 1.7 s	a= 544.1	Kcuentas tiene	69 líneas
Comp.104	m= 265.1471	Da	se(m)= 0.4 mDa	t=22.75 min	t=1364.7 s	se(t)= 2.6 s	a= 810.3	Kcuentas tiene	76 líneas
Comp.105	m= 433.2356	Da	se(m)= 0.4 mDa	t=22.81 min	t=1368.7 s	se(t)= 6.9 s	a= 1503.4	Kcuentas tiene	88 líneas
Comp.106	m= 265.1473	Da	se(m)= 0.3 mDa	t=22.90 min	t=1373.8 s	se(t)= 2.6 s	a= 832.6	Kcuentas tiene	116 líneas
Comp.107	m= 391.1400	Da	se(m)= 0.7 mDa	t=22.91 min	t=1374.9 s	se(t)=17.9 s	a= 2284.9	Kcuentas tiene	12 líneas
Comp.108	m= 471.3489	Da	se(m)= 0.8 mDa	t=23.04 min	t=1382.3 s	se(t)=12.9 s	a= 10293.8	Kcuentas tiene	30 líneas
Comp.109	m= 116.9276	Da	se(m)= 0.5 mDa	t=23.15 min	t=1388.7 s	se(t)= 7.7 s	a= 615.2	Kcuentas tiene	77 líneas
Comp.110	m= 311.1678	Da	se(m)= 0.5 mDa	t=23.20 min	t=1392.0 s	se(t)=23.2 s	a= 6878.2	Kcuentas tiene	9 líneas
Comp.111	m= 361.1291	Da	se(m)= 0.5 mDa	t=23.29 min	t=1397.5 s	se(t)= 5.4 s	a= 3463.5	Kcuentas tiene	166 líneas
Comp.112	m= 309.1726	Da	se(m)= 0.6 mDa	t=23.40 min	t=1404.2 s	se(t)=14.3 s	a= 1622.6	Kcuentas tiene	26 líneas
Comp.113	m= 116.9274	Da	se(m)= 0.3 mDa	t=23.43 min	t=1406.0 s	se(t)= 2.6 s	a= 237.6	Kcuentas tiene	52 líneas
Comp.114	m= 293.1775	Da	se(m)= 0.5 mDa	t=23.53 min	t=1411.6 s	se(t)= 4.0 s	a= 348.6	Kcuentas tiene	99 líneas
Comp.115	m= 340.2018	Da	se(m)= 0.9 mDa	t=23.53 min	t=1411.9 s	se(t)=15.2 s	a= 844.9	Kcuentas tiene	8 líneas
Comp.116	m= 433.2351	Da	se(m)= 0.4 mDa	t=23.64 min	t=1418.7 s	se(t)=18.0 s	a= 8889.4	Kcuentas tiene	10 líneas
Comp.117	m= 265.1464	Da	se(m)= 0.4 mDa	t=23.66 min	t=1419.4 s	se(t)= 5.2 s	a= 2010.8	Kcuentas tiene	178 líneas
Comp.118	m= 353.1996	Da	se(m)= 0.6 mDa	t=23.78 min	t=1426.9 s	se(t)= 7.8 s	a= 712.4	Kcuentas tiene	43 líneas
Comp.119	m= 312.1703	Da	se(m)= 0.4 mDa	t=23.78 min	t=1427.0 s	se(t)= 2.5 s	a= 216.6	Kcuentas tiene	26 líneas
Comp.120	m= 339.1991	Da	se(m)= 0.6 mDa	t=23.78 min	t=1427.1 s	se(t)=25.3 s	a= 7701.5	Kcuentas tiene	6 líneas
Comp.121	m= 311.1673	Da	se(m)= 0.3 mDa	t=23.88 min	t=1432.5 s	se(t)= 3.5 s	a= 1540.6	Kcuentas tiene	180 líneas
Comp.122	m= 325.1826	Da	se(m)= 0.4 mDa	t=23.98 min	t=1439.0 s	se(t)=11.2 s	a= 4681.4	Kcuentas tiene	50 líneas
Comp.123	m= 297.1515	Da	se(m)= 0.5 mDa	t=24.01 min	t=1440.4 s	se(t)= 4.3 s	a= 230.8	Kcuentas tiene	40 líneas
Comp.124	m= 293.1773	Da	se(m)= 0.7 mDa	t=24.10 min	t=1445.9 s	se(t)= 4.9 s	a= 764.6	Kcuentas tiene	128 líneas
Comp.125	m= 433.2358	Da	se(m)= 0.3 mDa	t=24.21 min	t=1452.6 s	se(t)= 3.5 s	a= 2183.4	Kcuentas tiene	192 líneas
Comp.126	m= 265.1465	Da	se(m)= 0.4 mDa	t=24.28 min	t=1456.6 s	se(t)=17.8 s	a= 5404.3	Kcuentas tiene	23 líneas
Comp.127	m= 433.2356	Da	se(m)= 0.4 mDa	t=24.35 min	t=1461.1 s	se(t)= 1.4 s	a= 897.8	Kcuentas tiene	99 líneas



Comp.128	m= 116.9275	Da	se (m)= 0.2	mDa	t=24.38	min	t=1463.0	s	se (t)= 4.3	s	a= 590.6	Kcuentas tiene	43	líneas
Comp.129	m= 227.2007	Da	se (m)= 0.5	mDa	t=24.44	min	t=1466.7	s	se (t)=10.9	s	a= 4481.8	Kcuentas tiene	48	líneas
Comp.130	m= 311.1680	Da	se (m)= 0.3	mDa	t=24.48	min	t=1468.6	s	se (t)= 1.7	s	a= 724.2	Kcuentas tiene	96	líneas
Comp.131	m= 326.1864	Da	se (m)= 0.8	mDa	t=24.48	min	t=1469.0	s	se (t)= 4.3	s	a= 454.9	Kcuentas tiene	8	líneas
Comp.132	m= 293.1772	Da	se (m)= 0.5	mDa	t=24.57	min	t=1474.1	s	se (t)= 2.5	s	a= 355.6	Kcuentas tiene	69	líneas
Comp.133	m= 309.1749	Da	se (m)= 1.6	mDa	t=24.59	min	t=1475.5	s	se (t)= 9.5	s	a= 1313.5	Kcuentas tiene	56	líneas
Comp.134	m= 433.2355	Da	se (m)= 0.5	mDa	t=24.68	min	t=1480.8	s	se (t)=10.7	s	a= 5451.3	Kcuentas tiene	44	líneas
Comp.135	m= 367.2633	Da	se (m)= 1.5	mDa	t=24.79	min	t=1487.1	s	se (t)= 9.3	s	a= 5546.7	Kcuentas tiene	92	líneas
Comp.136	m= 325.1831	Da	se (m)= 0.5	mDa	t=24.88	min	t=1493.1	s	se (t)= 4.9	s	a= 1945.1	Kcuentas tiene	188	líneas



Compuestos obtenidos mediante el *clustering* de líneas y 10 líneas de mayor área de la huella espectrométrica de cada uno. Se presentan únicamente aquellos que habían sido previamente identificados en la caracterización convencional de aceite:

Comp	4	m=123.0444	Da	t=6.5 min	a=1241.0 Kcuentas
	123.0444	6.5	1240.95200	100.0000	
	153.0555	6.5	1192.86400	96.1249	
	122.0364	6.5	109.31200	8.8087	
	154.0587	6.5	109.31200	8.8087	
	124.0476	6.5	101.27200	8.1608	
	251.0222	6.5	56.50800	4.5536	
	235.0457	6.5	33.23200	2.6779	
	95.0464	6.5	22.50800	1.8138	
	326.0083	6.5	21.47600	1.7306	
	297.0515	6.5	21.12800	1.7026	
Comp	5	m=137.0595	Da	t=8.2 min	a=210.7 Kcuentas
	137.0595	8.2	210.68400	100.0000	
	263.0333	8.1	60.86400	28.8888	
	374.9909	8.2	35.72400	16.9562	
	221.0218	8.2	26.87200	12.7546	
	240.9677	8.1	24.87200	11.8054	
	106.0407	8.2	22.14400	10.5105	
	138.0638	8.2	19.56000	9.2840	
	359.0134	8.1	18.87600	8.9594	
	119.0495	8.2	16.44400	7.8051	
	341.0861	8.2	14.44800	6.8577	
Comp	24	m=335.1136	Da	t=13.9 min	a=676.7 Kcuentas
	335.1136	13.9	676.66000	100.0000	
	377.1235	13.8	475.53600	70.2769	
	307.0812	13.8	274.00400	40.4936	
	184.0692	13.9	246.27200	36.3952	
	199.0607	13.9	240.02400	35.4719	
	95.0467	13.9	211.22400	31.2157	
	213.0759	13.8	165.02400	24.3880	
	336.1163	13.9	123.24800	18.2142	
	378.1276	13.9	106.40400	15.7249	
	113.0235	14.0	91.24800	13.4851	
Comp	26	m=377.1231	Da	t=14.2 min	a=1948.1 Kcuentas
	377.1231	14.2	1948.08800	100.0000	
	275.0892	14.2	1302.69600	66.8705	
	307.0809	14.2	991.25200	50.8833	
	349.1287	14.2	511.22000	26.2421	
	378.1268	14.2	476.33200	24.4513	
	139.0448	14.2	453.92400	23.3010	
	149.0236	14.2	355.44400	18.2458	
	213.0758	14.1	309.09200	15.8664	
	345.0965	14.2	290.19600	14.8965	
	399.1048	14.2	273.96000	14.0630	
Comp	29	m=319.1184	Da	t=14.5 min	a=12193.0 Kcuentas
	319.1184	14.5	12192.98800	100.0000	
	639.2443	14.5	5790.08000	47.4870	
	640.2477	14.5	2335.72000	19.1563	
	641.2497	14.5	533.05600	4.3718	
	401.1162	14.5	231.07200	1.8951	
	325.1825	14.5	77.54400	0.6360	



663.2126	14.5	74.21200	0.6086	
362.1302	14.4	50.06400	0.4106	
127.0381	14.4	43.35200	0.3555	
259.0959	14.5	42.28400	0.3468	
Comp 31	m=377.1247	Da	t=14.8 min	a=4665.2 Kcuentas
377.1247	14.8	4665.24000	100.0000	
307.0816	14.9	1798.86800	38.5590	
275.0897	14.8	1614.81600	34.6138	
378.1278	14.8	970.90800	20.8115	
349.1295	14.9	663.72000	14.2269	
320.1219	14.9	655.42000	14.0490	
213.0762	14.9	593.34000	12.7183	
149.0240	14.8	563.10000	12.0701	
139.0418	14.8	562.12800	12.0493	
345.0975	14.8	515.13600	11.0420	
Comp 41	m=285.0401	Da	t=16.0 min	a=1296.0 Kcuentas
285.0401	16.0	1296.00400	100.0000	
275.0892	16.0	234.37600	18.0845	
286.0428	16.0	219.75600	16.9564	
363.1438	16.0	136.52400	10.5342	
311.1680	15.9	131.94400	10.1808	
321.1008	16.0	126.37200	9.7509	
139.0409	15.9	109.61200	8.4577	
361.1299	15.9	89.09200	6.8744	
227.0918	16.0	87.38800	6.7429	
266.1501	16.0	63.88000	4.9290	
Comp 42	m=391.1402	Da	t=16.1 min	a=316.2 Kcuentas
391.1402	16.1	316.24000	100.0000	
Comp 44	m=377.1249	Da	t=16.3 min	a=2260.4 Kcuentas
377.1249	16.3	2260.44000	100.0000	
307.0818	16.2	826.76800	36.5755	
275.0896	16.3	700.68800	30.9979	
378.1284	16.3	461.48800	20.4158	
345.0980	16.2	287.53200	12.7202	
149.0243	16.2	269.85600	11.9382	
139.0369	16.2	256.12400	11.3307	
319.1181	16.2	200.46000	8.8682	
95.0469	16.2	173.73600	7.6859	
399.1060	16.2	148.46400	6.5679	
Comp 47	m=361.1294	Da	t=16.6 min	a=1250.6 Kcuentas
361.1294	16.6	1250.58400	100.0000	
291.0864	16.5	1019.29600	81.5056	
213.0761	16.6	808.89200	64.6811	
377.1242	16.6	494.34800	39.5294	
259.0966	16.6	343.94800	27.5030	
333.1339	16.5	238.82400	19.0970	
362.1324	16.6	228.60400	18.2798	
127.0388	16.6	208.55600	16.6767	
311.1676	16.5	126.66800	10.1287	
383.1117	16.5	97.55200	7.8005	
Comp 48	m=265.1473	Da	t=16.7 min	a=515.8 Kcuentas
265.1473	16.7	515.82400	100.0000	
391.1402	16.7	395.42000	76.6579	
275.0892	16.7	374.55600	72.6131	
357.1341	16.7	326.12800	63.2247	
321.0980	16.7	224.12800	43.4505	
149.0245	16.7	223.77200	43.3815	
139.0391	16.7	211.68400	41.0380	
165.0555	16.7	211.28400	40.9605	



95.0466	16.7	192.71200	37.3600		
378.1284	16.7	166.56800	32.2916		
Comp 52	m=183.0659	Da	t=17.3 min	a=420.9	Kcuentas
183.0659	17.3	420.92400	100.0000		
265.1473	17.3	287.32800	68.2613		
393.1201	17.4	247.04400	58.6909		
165.0550	17.3	184.95600	43.9405		
325.1821	17.3	158.75200	37.7151		
233.0808	17.3	136.51200	32.4315		
311.1676	17.4	93.84400	22.2948		
285.1121	17.3	79.11600	18.7958		
123.0445	17.3	71.91600	17.0853		
415.1276	17.3	66.43600	15.7834		
Comp 55	m=361.1290	Da	t=17.5 min	a=2092.2	Kcuentas
361.1290	17.5	2092.19200	100.0000		
291.0860	17.5	1786.28400	85.3786		
213.0757	17.5	1742.46800	83.2843		
377.1238	17.5	1674.91200	80.0554		
307.0808	17.5	792.61600	37.8845		
275.0886	17.5	615.58800	29.4231		
259.0961	17.5	585.94000	28.0060		
95.0465	17.4	554.83200	26.5192		
333.1334	17.5	544.12800	26.0076		
111.0077	17.5	481.11200	22.9956		
Comp 56	m=183.0653	Da	t=17.7 min	a=801.8	Kcuentas
183.0653	17.7	801.77200	100.0000		
265.1460	17.6	515.96800	64.3535		
303.1222	17.7	295.74400	36.8863		
311.1675	17.7	259.38800	32.3518		
165.0548	17.7	257.45600	32.1109		
233.0800	17.7	197.62400	24.6484		
285.1112	17.7	193.33200	24.1131		
123.0439	17.7	148.65600	18.5409		
325.1815	17.7	127.06000	15.8474		
339.1986	17.7	95.86400	11.9565		
Comp 58	m=377.1242	Da	t=17.9 min	a=3825.1	Kcuentas
377.1242	17.9	3825.05600	100.0000		
307.0811	17.9	1571.80400	41.0923		
275.0890	17.9	1331.14400	34.8006		
378.1275	17.9	734.40800	19.1999		
139.0379	17.9	534.54800	13.9749		
149.0237	17.9	482.55600	12.6157		
345.0970	17.9	455.73600	11.9145		
95.0466	17.9	369.66800	9.6644		
183.0658	18.0	356.60400	9.3228		
308.0846	17.9	292.26000	7.6407		
Comp 65	m=265.1469	Da	t=18.7 min	a=1036.9	Kcuentas
265.1469	18.7	1036.91600	100.0000		
183.0657	18.7	472.96400	45.6126		
269.0444	18.7	270.46400	26.0835		
291.0865	18.7	212.65200	20.5081		
361.1296	18.7	207.96800	20.0564		
233.0802	18.6	169.88000	16.3832		
259.0965	18.7	119.44400	11.5192		
339.1990	18.8	109.98800	10.6072		
433.2353	18.7	99.81200	9.6259		
285.1116	18.7	69.70800	6.7226		
Comp 67	m=377.1245	Da	t=18.9 min	a=6055.1	Kcuentas
377.1245	18.9	6055.11600	100.0000		



307.0812	18.9	2397.69600	39.5979	
275.0895	18.9	2174.23200	35.9074	
378.1278	19.0	1495.50800	24.6983	
149.0238	19.0	872.94000	14.4166	
345.0973	18.9	808.94000	13.3596	
139.0365	18.9	585.41600	9.6681	
95.0467	19.0	569.82800	9.4107	
755.2555	18.9	516.27200	8.5262	
399.1060	19.0	482.33200	7.9657	
Comp 68	m=265.1471	Da	t=19.2 min	a=1256.2 Kcuentas
265.1471	19.2	1256.20800	100.0000	
361.1289	19.1	465.40400	37.0483	
291.0859	19.1	324.02400	25.7938	
183.0656	19.2	319.46000	25.4305	
127.0388	19.2	143.65200	11.4354	
242.1754	19.2	137.91200	10.9784	
325.1825	19.2	114.00400	9.0752	
362.1320	19.1	113.66400	9.0482	
276.0925	19.1	112.54400	8.9590	
339.2000	19.2	101.42800	8.0741	
Comp 69	m=275.0889	Da	t=19.3 min	a=375.5 Kcuentas
275.0889	19.3	375.46000	100.0000	
307.0808	19.2	313.95600	83.6190	
345.0972	19.4	255.14000	67.9540	
433.2356	19.4	192.32400	51.2236	
241.0706	19.3	160.48400	42.7433	
111.0077	19.3	132.13600	35.1931	
311.1680	19.3	127.86800	34.0564	
95.0467	19.4	103.26800	27.5044	
299.0542	19.3	91.54400	24.3818	
325.1827	19.4	86.84800	23.1311	
Comp 74	m=377.1246	Da	t=19.7 min	a=7197.3 Kcuentas
377.1246	19.7	7197.34400	100.0000	
101.0217	19.7	101.45600	1.4096	
147.0444	19.7	33.06800	0.4594	
329.2318	19.7	30.42400	0.4227	
277.0936	19.6	25.44400	0.3535	
99.0410	19.7	10.00400	0.1390	
335.1130	19.7	5.94000	0.0825	
341.1142	19.7	5.65600	0.0786	
294.1822	19.7	5.25200	0.0730	
398.2299	19.7	4.94400	0.0687	
Comp 77	m=265.1474	Da	t=19.9 min	a=533.3 Kcuentas
265.1474	19.9	533.25200	100.0000	
275.0898	19.9	470.45600	88.2240	
361.1297	19.9	202.93200	38.0556	
756.2593	19.9	161.37600	30.2626	
111.0082	20.0	143.33600	26.8796	
311.1684	20.0	143.29600	26.8721	
379.1306	19.9	129.66000	24.3150	
433.2358	19.9	127.16800	23.8476	
95.0466	20.0	124.68000	23.3811	
346.1012	19.9	119.57600	22.4239	
Comp 83	m=361.1299	Da	t=20.6 min	a=1399.9 Kcuentas
361.1299	20.6	1399.91200	100.0000	
291.0868	20.6	1099.70400	78.5552	
362.1337	20.6	268.21600	19.1595	
339.2002	20.7	211.96000	15.1410	
325.1835	20.7	191.77200	13.6989	



292.0903	20.6	173.98800	12.4285	
459.1192	20.7	116.93600	8.3531	
99.9235	20.6	85.20000	6.0861	
535.1130	20.7	80.08400	5.7206	
312.1715	20.6	66.44000	4.7460	
Comp 85	m=377.1263	Da	t=20.8 min	a=26611.4 Kcuentas
377.1263	20.8	26611.38000	100.0000	
275.0898	20.8	6686.70800	25.1273	
755.2572	20.8	6565.59600	24.6721	
149.0240	20.8	2950.12400	11.0859	
756.2594	20.8	2574.50000	9.6744	
139.0362	20.8	2362.42000	8.8775	
95.0468	20.8	1849.07600	6.9484	
265.1475	20.7	1258.36400	4.7287	
111.0081	20.8	1218.88000	4.5803	
379.1301	20.8	924.94800	3.4758	
Comp 93	m=377.1242	Da	t=21.6 min	a=2112.5 Kcuentas
377.1242	21.6	2112.50000	100.0000	
307.0813	21.6	882.68800	41.7840	
275.0896	21.6	698.86000	33.0821	
378.1279	21.6	410.26000	19.4206	
149.0238	21.6	390.36000	18.4786	
325.1830	21.5	373.63200	17.6867	
345.0975	21.6	361.76800	17.1251	
433.2354	21.5	275.60400	13.0463	
139.0353	21.6	245.72000	11.6317	
95.0466	21.6	206.88800	9.7935	
Comp 96	m=377.1240	Da	t=21.9 min	a=954.6 Kcuentas
377.1240	21.9	954.62400	100.0000	
361.1291	21.9	656.81600	68.8036	
291.0864	22.0	451.62400	47.3091	
275.0897	21.9	334.53600	35.0437	
307.0814	21.9	260.32000	27.2694	
116.9276	21.9	247.21600	25.8967	
378.1277	21.9	195.86800	20.5178	
297.1516	21.9	177.86000	18.6314	
149.0239	21.9	150.09200	15.7226	
345.0974	22.0	144.38800	15.1251	
Comp 107	m=391.1400	Da	t=22.9 min	a=2284.9 Kcuentas
391.1400	22.9	2284.94800	100.0000	
473.3540	22.8	300.39600	13.1467	
553.3503	22.9	281.50800	12.3201	
752.6198	22.9	260.16000	11.3858	
753.6216	22.9	147.78000	6.4675	
554.3533	22.9	91.44800	4.0022	
629.3366	22.9	72.66000	3.1799	
569.3249	22.9	52.16400	2.2829	
583.3587	22.9	50.23200	2.1984	
515.2372	22.8	48.84800	2.1378	
Comp 111	m=361.1291	Da	t=23.3 min	a=3463.5 Kcuentas
361.1291	23.3	3463.48000	100.0000	
265.1468	23.3	3367.49200	97.2286	
291.0860	23.3	1995.52800	57.6163	
362.1319	23.3	682.38000	19.7021	
312.1709	23.3	558.42800	16.1233	
127.0391	23.3	553.27200	15.9745	
326.1865	23.3	479.79600	13.8530	
101.0216	23.3	400.55600	11.5651	
259.0961	23.3	344.88000	9.9576	



241.0705	23.3	313.83600	9.0613	
Comp 116	m=433.2351	Da	t=23.6 min	a=8889.4 Kcuentas
433.2351	23.6	8889.42400	100.0000	
434.2388	23.6	2030.82400	22.8454	
361.1285	23.7	471.50400	5.3041	
151.0396	23.7	135.67200	1.5262	
101.0214	23.7	79.72000	0.8968	
397.3639	23.6	47.17600	0.5307	
541.2712	23.7	7.07200	0.0796	
221.0838	23.6	7.02400	0.0790	
663.3242	23.7	5.77200	0.0649	
758.2344	23.7	5.10800	0.0575	
Comp 129	m=227.2007	Da	t=24.4 min	a=4481.8 Kcuentas
227.2007	24.4	4481.78800	100.0000	
325.1829	24.5	3464.05600	77.2918	
228.2032	24.4	784.94800	17.5142	
383.3520	24.4	177.94000	3.9703	
417.3180	24.5	124.56000	2.7792	
525.2706	24.5	77.46400	1.7284	
585.5054	24.5	76.74800	1.7124	
229.2058	24.4	75.18800	1.6776	
579.2848	24.5	58.02800	1.2948	
481.2570	24.5	52.04400	1.1612	



