**OPEN ACCESS DOCUMENT**

Information of the Journal in which the present paper is published:

* J. of Chromatography, 2018
* DOI: <https://doi.org/10.1016/j.chroma.2018.07.017>

**An untargeted lipidomic strategy combining comprehensive two-dimensional liquid chromatography and chemometric analysis**

Meritxell Navarro-Reig, Joaquim Jaumot\* and Romà Tauler

Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain

Corresponding author: Joaquim Jaumot

Postal address: Department of Environmental Chemistry, IDAEA-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain.

Telephone: 934006100 - 1443

E-mail: joaquim.jaumot@idaea.csic.es

# **Abstract**

Untargeted lipidomic samples are extremely complex and often exceed the limits of peak capacity achievable by one-dimensional liquid chromatography (LC). Comprehensive two-dimensional liquid chromatography (LC×LC) appears as a promising alternative to overcome this drawback. Unfortunately, this approach generates highly complex datasets which untargeted analysis is challenging. In this work, a global methodological strategy combining LC×LC-MS with chemometric data analysis is proposed for untargeted lipidomic studies. The feasibility of the proposed methodology is demonstrated by its application to assess the effects of arsenic exposure on the lipidome of growing rice samples. A two-dimensional chromatographic setup coupling reversed phase (RP) and hydrophilic interaction liquid chromatography (HILIC) modes together with a triple quadrupole mass detector (TQD) is proposed to analyze lipid extracts from rice samples at different experimental conditions. Chemometric tools were used for data compression, spectral and elution profiles resolution, feature detection and statistical analysis of the multidimensional LC×LC data. The obtained results revealed that the proposed methodology was useful to gather relevant information from untargeted lipidomic studies and detect potential biomarkers.

# **Keywords**

Untargeted lipidomics, LC×LC-MS, comprehensive, chemometrics, ROIMCR.

# **Introduction**

Lipidomics is the branch of metabolomics consisting in the comprehensive study of lipid species and their related networks and metabolomic pathways of a biological system. Lipids are a group of biomolecules involved in a wide range of structural and functional activities in cells. For example, lipids are structural components of cell membranes influencing both, its fluidity and its interactions (such as nutrient transport or waste products expulsion). Moreover, lipids are also involved in energy transport and storage, as well as in cell communication and signaling. For this reason, the comprehensive analysis of lipids is gaining more attention in many research fields, from biomedical to environmental studies [1-5].

There are two major analytical strategies in lipidomic studies: targeted and untargeted approaches. The target approach focuses on analyzing a specific list of lipids, typically related to some known pathway of interest, in an attempt to corroborate a previous hypothesis. In contrast, untargeted lipidomics aims to the global analysis of all measurable lipids present in a sample, without any prior assumption about affected pathways or lipid species [4, 6-8]. In this work, the untargeted approach is used, with the aim of discovering which lipid species are associated with previously unexplored biological pathways [9].

The main drawback of untargeted lipidomics is the high complexity of the generated data because of the structural diversity of lipids. The International Lipid Classification and Nomenclature Committee established a “Comprehensive Classification System for Lipids” based on their chemical and biochemical properties [10]. This classification system grouped lipids into eight different categories: fatty acids, glycerolipids, glycerophospholipids, sphingolipids, sterols, prenol lipids, saccharolipids, and polyketides. Moreover, there are several subcategories within each of the mentioned groups, resulting in a large number of combinations [1]. This great structural diversity and chemical complexity among lipids give a wide range of different physical properties, which causes the profiling of the complete lipidome of biological samples to be still a challenge [1-3, 11].

To overcome this hurdle, analytical platforms used in untargeted lipidomics must have a high separation power and at the same time be highly sensitive and selective. Until now, liquid chromatography (LC) coupled to mass spectrometry (MS) is the most frequently used analytical platform in lipidomic studies [1, 2, 6-8, 11]. Nevertheless, biological lipid samples contain thousands of lipids and many of them severely coeluted. The high complexity of these biological lipid extracts often exceeds the limits of peak capacity achievable by LC systems. Consequently, despite its high capacity resolving complex samples, LC-MS may still lose important information in lipidomic studies. Therefore, the use of multidimensional separation systems coupled to MS is proposed to overcome this drawback.

Among multidimensional analytical platforms, comprehensive two-dimensional liquid chromatography (LC×LC) offers two relevant advantages for untargeted lipidomic studies. The first benefit is related to the increase of the resolution capacity in comparison with mono-dimensional systems. The higher resolution power of LC×LC lies in the fact that under ideal circumstances (when the two separations systems are completely orthogonal), the total peak capacity is equal to the product of individual peak capacities of the first- and the second-dimension separations. In practice, the complete orthogonality of both separation systems is difficult to achieve, but a high peak capacity can be reached if two uncorrelated separation modes are used, such as reverse phase (RP) and hydrophilic interaction liquid chromatography (HILIC) [12-14]. The other significant advantage of LC×LC is that this powerful approach increases the identification ability because of the information provided on the different separation patterns of the two columns. In 2D chromatograms, peaks are usually observed along lines or arcs related to specific analyte functionalities (alkanes, aldehydes or degree of unsaturation among others), which provides additional help for identifying unknown lipids [12].

In contrast, the main drawback of LC×LC-MS is that generates complex datasets where the relevant information can remain hidden. Thousands of signals can be detected when analyzing lipid extracts by LC×LC-MS. The manual inspection of these signals is not feasible in practice and, therefore, their processing is not straightforward [15]. In order to achieve a complete resolution of complex lipid extracts and gather as much information as possible from the analyzed biological system, the use of advanced chemometric methods is recommended [15-17]. However, in the analytical literature, only a few number of works have been focused on the chemometric analysis of LC×LC data. Also, up to date, most of these works deal only with LC×LC-DAD (diode array detector) data [14, 15, 17-19].

The primary goal of this work is to present a new global methodological strategy combining the LC×LC-MS powerful analytical approach with advanced chemometric data analysis tools. This procedure will provide as much information as possible from untargeted lipidomic studies. With this aim, the combination of the LC×LC-MS/MS method (coupling RPLC with HILIC) with some chemometric tools is proposed for the analysis of the entire lipidome of complex biological samples. The feasibility of the proposed approach is demonstrated by its application to the study of the changes produced on Japanese rice (*Oryza sativa* L.) lipidome under arsenic exposure.

# **2. Materials and Methods**

## 2.1. Chemicals and Reagents

Sodium arsenate dibasic heptahydrate (≥98.0 %), HPLC grade water, HPLC grade acetonitrile, HPLC grade isopropanol, methanol (MeOH, HPLC grade), methyl tert-butyl ether (MTBE), ammonium acetate (≥99.0%), acetic acid (≥95.0%) and formic acid (≥95.0%) were supplied by Sigma-Aldrich (Steinheim, Germany).

Eight lipid standards from different families were used as extraction standards: 17:0 monoacylglycerol, 17:1 lysophosphatidylethanolamine, 17:0 lysophosphatidylcholine, 1,3-17:0 D5 diacylglyceride, 17:0 cholesteryl ester, 1,2,3-17:0 triglyceride, 16:0 D31-18:1 phosphatidylcholine, 16:0 D31-18:1 phosphatidylserine. Three sphingolipids were used as internal standards: N-dodecanoylsphingosine, N-dodecanoylglucosyl-sphingosine and N-dodecanoylsphingosylphosphorylcholine. All these lipid standards were obtained from Avanti Polar Lipids (Alabaster, AL, USA).

Solutions containing 1 and 1000 µM of arsenic (As) were weekly prepared by adequately diluting a 10000 µM stock solution. The stock solution was prepared by dissolution of the appropriate amount of sodium arsenate salt. All solutions were stored at 6 ºC until their use.

Water used for plant watering and for preparing arsenic solutions was purified using an Elix 3 coupled to a Milli-Q system (Millipore, Belford, MA, USA), and filtered through a 0.22 µm nylon filter integrated into the Milli-Q system.

The following abbreviations have been used to describe lipid families: (PA) phosphatidic acid, (PC) phosphatidylcholines, (PG) Phosphatidylglycerol, (PI) phosphatidylinositols, (Cer) Ceramides, (MGDG) monogalactosyldiacylglycerol, (DAG) diacylglycerols, (TAG) triacylglycerols.

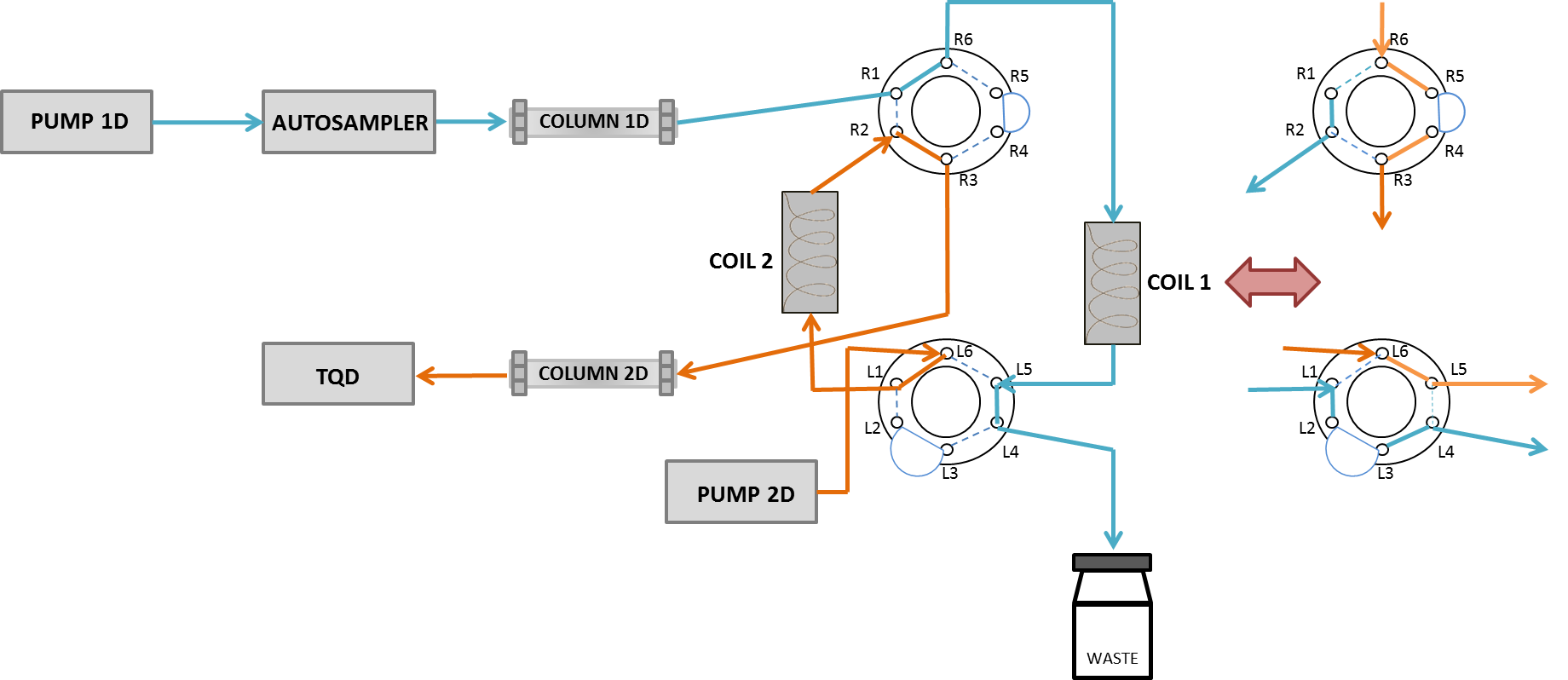
## 2.2. Plant growth and sample preparation

Plant growth and lipid extraction were performed using the procedure described elsewhere [20, 21]. Briefly, rice seeds, obtained from the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain), were incubated for two days at 30 ºC in a wet environment. After this period, plants were grown on an Environmental Test Chamber MLE-352H (Panasonic®) for 22 days simulating cyclic environmental changes of temperature, relative humidity, and light intensity, as shown in the Supplementary Information (Figure S1). During the first 10 days of growth, rice plants were watered with Milli-Q water three times per week. Since then, the plant treated samples were subjected to irrigation water containing low (1 µM) or high (1000 µM) concentrations of As, whereas the plant control samples were watered with Milli-Q water until harvest. The lower concentration was set to 1 µM because it is the limit of the acceptable As concentration in water by European legislation (Groundwater Directive2006/118/EC) [22]. After harvest, aerial parts (part of the plant above ground) and roots were separated and immediately samples were frozen at liquid nitrogen temperature for metabolism quenching. Then, samples were stored at -80 ºC until extraction. Five biological replicates were made for each sample condition. Therefore, a total number of 30 samples (15 aerial part samples and 15 root samples) were analyzed.

Before extraction, rice samples were ground to a fine powder using a liquid nitrogen mortar and lyophilized for 24 hours to dryness. Lipid extraction was carried out by dispersing 10 mg of the dried tissue in 1 mL of MTBE:MeOH (3:1). The mixture was fortified with 20 µL of the extraction standards mix, and then, vortexed for 1 min and sonicated for 10 min. Next, 0.5 mL of H2O:MeOH (3:1) were added, and the mixture was again vortexed for 1 min. After centrifuging for 5 min at 2000 x g, the organic fraction (upper) was collected. The aqueous phase (lower) was re-extracted with 0.65 mL of MTBE and 0.35 mL of MeOH:H2O (1:0.85). Next, the mixture was vortexed for 1 min and centrifuged for 5 min at 2000 x g. After that, organic phases were combined and evaporated to dryness under nitrogen gas. All of the extracts were stored at -80ºC until analysed. Before injection, extracts were reconstituted with 250 mL of MeOH:H2O (4:1) and 20 µL of the internal standards mix were added. Quality control (QC) samples were prepared by pooling 15 µL of all studied samples (extracts). QC for aerial part and root samples were separately prepared.

## 2.3.LC×LC-MS/MS analysis

LC×LC analyses were carried out on an Acquity UHPLC system (Waters, Milford, MA, USA) equipped with a quaternary pump and an autosampler. Second-dimension separation was possible due to the coupling to this instrument of one additional LC pump (Waters 1525 binary HPLC pump). The interface between the first and the second column was an Acquity UPLC Column Manager (Waters, Milford, MA, USA), equipped with two 6-port two-position valves. Figure 1 shows the scheme of the LC×LC system used in this work.



**Figure 1.** Scheme of the LC×LC-MS system. The red double arrow shows the change in modulator.

In the first chromatographic dimension, an RP ZORBAX Eclipse XDB-C18 (150 mm × 2.1 mm i.d.; 5µm) column provided by Agilent (Santa Clara, CA, USA) was used. Chromatographic analysis was run using (A) acetonitrile:isopropanol (1:2) 0.1% formic acid and (B) water 0.1% formic acid, as mobile phase, eluted according to the following gradient: 0 min, 80% A; 20.5 min 90% A; 78.0-98.5 min, 100% A; 98.5-99.5 min, back to initial conditions at 80% A and from 99.5 to 130 min, at 80% A. The mobile phase flow rate was 39 µL·min-1, and the injection volume was 20 µL.

The second chromatographic dimension employed a KINETEX HILIC (30 mm × 3 mm i.d.; 2.6 µm) column provided by Phenomenex (Torrance, CA, USA). Second-dimension elution gradient used 5 mM ammonium acetate at pH 5.5, adjusted with acetic acid (A) and acetonitrile (B), in an isocratic elution gradient at 16% A. The modulation time in the switching valve was set at 1.8 min. The mobile phase flow rate was 0.5 mL·min-1.

Mass spectroscopic detection was performed in a triple quadrupole detector (TQD, Waters, Milford, MA, USA) equipped with an electrospray (ESI) as ionization source working in both negative and positives modes. Nitrogen (purity ˃99.98 %) was used as desolvation gas at the flow rate of 800 L·h-1. Desolvation temperature was set at 450 ºC, and the con voltage was set at 50 V. First, samples were analysed in full scan mode in an untargeted manner using a mass acquisition range from 90 to 1800 Da. Then, after the application of the chemometric data analysis strategy (see the following section), the most important mass traces were selected. Finally, samples were reinjected in the same chromatographic conditions to obtain the MS/MS spectra of the selected mass traces. All MS/MS spectra were recorded at 10, 20, 30 and 40 eV collision energies (CE).

## 2.4. Chemometric data analysis strategy

The first step of the proposed data analysis strategy consisted on the compression and the arrangement of the raw LC×LC-MS data using an approach based on the selection of the so-called regions of interest (ROI) [23]. Data matrices generated by this strategy were then analyzed and resolved by means of the multivariate curve resolution by alternating least squares method (MCR-ALS) [24, 25], which provides the pure elution and mass spectra profiles of the constituents (lipids) present in the analyzed samples. The procedure that combines the ROI data compression and the MCR-ALS analysis of the ROI compressed data is called ROIMCR and has been described in more detail elsewhere [23]. A further data analysis step was the statistical assessment of the effects of As exposure on rice based on the changes of the peak areas of the elution profiles of the resolved lipidic constituents by principal component analysis (PCA) [26] and ANOVA-simultaneous component analysis (ASCA) [27]. In addition, partial least squares – discriminant analysis (PLS-DA) [28] allowed the differentiation between control and As treated samples and the identification of the main features that allowed this differentiation. Lipids whose concentration changed because of As treatment were tentatively identified using their MS/MS spectra. In the following sections, these different steps are described in more detail.

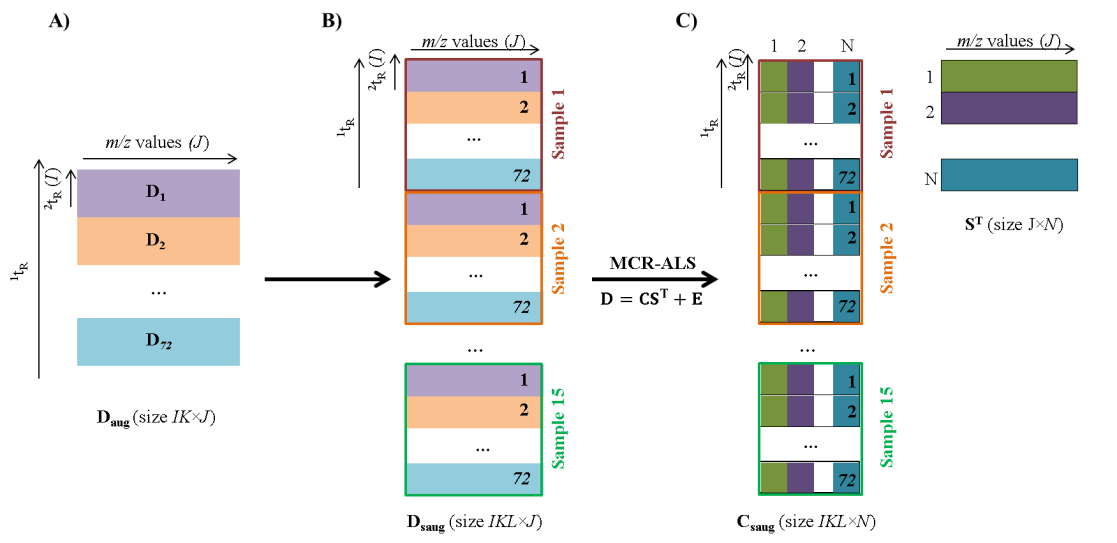
### 2.4.1. Compression and data arrangement

Waters raw chromatographic data files (.raw format) were converted to the standard CDF format by the Databridge function of MassLynxTM 4.1 software (Waters, Milford, MA, USA). Then, these data files were imported into MATLAB environment (Release 2016b, The Mathworks Inc, Natick, MA, USA) by using mzcdfread.m and mzcdf2peak.m functions of the MATLAB Bioinformatics Toolbox (4.3.1 version).

ROI strategy [23] was used to compress the MS data without losing mass accuracy and to build the data matrices to be analyzed by the MCR-ALS method. This strategy allowed the selection of the most interesting mass traces, which means those *m/z* values whose intensity signals were higher than a fixed signal-to-noise ratio threshold (SNRThr) and appeared a minimum number of times in the time direction. The required parameters for the implementation of the ROI approach were the SNRThr (set at 0.1% of the maximum MS signal intensity of each sample), the mass accuracy of the spectrometer (set at 0.5 Da/e for the TQD analyzer used in this work) and the minimum number of times at which the signal is acquired to be considered as a chromatographic peak (set at 25). These ROI values are searched along the entire chromatogram. Using this approach, the data matrix containing the intensities at all retention times (rows) for the selected number of ROI *m/z* values (columns) was finally obtained for each sample. More details about ROI strategy can be found at the work of Gorrochategui [23] and in Supplementary Information. A total number of 60 ROI matrices were obtained, one for each of the 30 samples of the presented work acquired at both ionization modes (positive and negative). Every ROI matrix was then normalized to correct for possible instrumental intensity changes among different sample injections and unavoidable differences in sample handling. This normalization was done by diving each matrix by the mean of the chromatographic area of the seven extraction standards and the three internal standards added to the lipidomic extract of each sample.

Once ROI compression was performed, every full-scan LC×LC-MS chromatographic run was arranged in a two-way data structure, as shown in Figure 2A. For each second-dimension chromatographic modulation, an individual LC-MS data matrix (**Dk**, see figure 2A) was built. These **Dk** matrices contained the second-dimension times on the rows and the *m/z* values on the columns. When *K* modulations were considered simultaneously, an LC×LC-MS column-wise augmented data matrix (**Daug**, see figure 2A) was built up settling the individual **Dk** matrices from each modulation one on the top of each other, and keeping their *m/z* values in common. Thus, **Daug** matrix contained the retention times of all modulation (*K=1,…,72*) on the rows and the *m/z* values on the columns.

After normalization, individual ROI matrices for each sample were arranged in four single superaugmented data matrices (Figure 2B, **Dsaug**): one for the aerial part samples analyzed in positive mode (**DAP**); one for the aerial part samples analyzed in negative mode (**DAN**); one for the root samples analyzed in positive mode (**DRP**); and one for the root samples analyzed in negative mode (**DRN**). As individual data matrices had a different number of ROI *m/z* values; a preliminary search considering common and uncommon ROI *m/z* values among different data samples was performed. When in an individual compressed matrix a particular ROI *m/z* value did not exist, a low random intensity value at the noise level was assigned (below SNRThr). In this way, the final superaugmented data matrices had the same ROI *m/z* values for the 15 samples simultaneously analyzed. For more details about this ROI matrix augmentation strategy see the Supporting Information and the protocol described in the work of Gorrochategui *et al.* [23]. Next, every one of these four superaugmented data matrices was analyzed by the MCR-ALS method.



**Figure 2**. A) LC×LC-MS analysis of a single sample (run) arranged in a column-wise augmented data matrix (**Daug**). Every single **Dk** matrix corresponds to one LC-MS second column modulation. B) LC×LC-MS analysis of multiple samples (runs) arranged in a column-wise superaugmented data matrix (**Dsaug**), with all the analysed samples (chromatographic runs) and their corresponding second column modulations. C) MCR-ALS resolution of LC×LC-MS data. Matrix **Dsaug** is decomposed into two matrices: **Csaug** which has the resolved pure elution profiles of the *N* components in all second column modulations of different samples (chromatographic runs), and **ST**, which has the pure mass spectra of the corresponding resolved components.

### 2.4.2. MCR-ALS resolution of LC×LC-MS data.

MCR-ALS is a chemometric method that allows the resolution of pure contributions present in unresolved complex mixtures. In this work, MCR-ALS was applied for the resolution of the elution profiles of the lipid constituents of the samples in both chromatographic dimensions and of their pure mass spectra. The MCR-ALS method has been already extensively described in the literature [24, 25, 29] and is only briefly explained here focusing on the particular case of untargeted multisample LC×LC-MS data.

MCR-ALS decomposes the experimental data sets according to a bilinear model extended to the simultaneous analysis of a large number of chromatographic runs (multiple second column modulations from several samples). For instance, in the case of this work, each of the four column-wise superaugmented data matrices (**Dsaug**, Figure 2B) had information related to the 15 single sample augmented LC×LC-MS data matrices (**Daug**, Figure 2A). **Dsaug** matrices were decomposed by MCR-ALS method using a bilinear model as shown in Equation 1 and Figure 2C.

Equation (1)

MCR-ALS decomposition of matrix **Dsaug** (*IKL*×*J*) gave **Csaug** (*IKL*×*N*) and **ST** (*N*×*J*). **Csaug** (*IKL*×*N*) contained second-dimension resolved elution profiles of the *N* components at all retention times (*I=59*) for each modulation (*K=72*) and sample (*L=15*). From this **Csaug** matrix, chromatographic peak areas of the resolved profiles of the different lipids and their relative quantitative information in the 15 analyzed samples can be obtained. On the other hand, **ST** (*N*×*J*) contained the pure mass spectra of the resolved components. **Esaug** contains the residual error not explained by the model.

More details regarding the initialization and constraints for the MCR-ALS optimization can be found in the Supplementary Information and references therein. MCR-ALS analyses were carried out using the MCR-ALS 2.0 toolbox available at [www.mcrals.info](http://www.mcrals.info).

### 2.4.3. Statistical assessment of As effects on rice lipidome

Chromatographic peak areas of the resolved lipids (*N*, in the columns) in all samples (*L*, in the rows) were arranged in a new data matrix (**A**). In total, four peak area matrices were obtained: aerial part samples analyzed in positive mode (**AAP**); aerial part samples analyzed in negative mode (**AAN**); root samples analyzed in positive mode (**ARP**); and root samples analyzed in negative mode (**ARN**). These peak area matrices were analyzed using PCA [26] and ASCA [27] to evaluate the effects of As exposure on rice.

PCA compresses the information of the original variables into a smaller number of uncorrelated variables known as principal components [26]. The representation of these components both in samples (scores maps) and variables (loadings) modes are useful to explore and interpret the variance sources in the analysed data.

ASCA is an extension of the multivariate analysis of variance method that combines the power of ANOVA to separate variance sources with the advantages of simultaneous component analysis (SCA) for the modelling of the individual separate factor effects matrices. In this work, ASCA was applied to statistically assess the significance of As exposure by using a permutation test in which the null hypothesis (H0) assumes that there is no effect of the considered factor. More details regarding the statistical assessment of ASCA results by using a permutation test can be found at the work of Vis *et. al.* [30]. In this work 1000 permutations has been used. Experimental design allowed performing ASCA analysis to well-balanced peak area matrices (*i.e.,* same number of samples for each analyzed condition). For a more detailed description of the ASCA procedure and permutation tests to assess the significance of factors see the works of Smilde [31], Jansen [27] and Vis [32].

Data was autoscaled before the application of PCA and only mean-centered before applying ASCA. Both methods were applied using PLS Toolbox 8.0.2 (Eigenvector Research Inc, Wenatchee, WA, USA) working under MATLAB 2015b.

### 2.4.4 Feature detection

After statistical evaluation of As effects on rice lipidome, PLS-DA [28] was used to detect what variables (lipids) were responsible for the observed differences between control and As-treated samples.

PLS-DA is a supervised multivariate regression method oriented to discriminate among different groups of samples. In this work, PLS-DA discriminated between control and As-treated samples. Here, PLS-DA was used to correlate the matrix of peak areas (**A**, predictor variable) with the vector describing the sample type class membership (**y**, predicted variable) [33]. Apart from sample class discrimination, PLS-DA also provides information about which are the most relevant variables for achieving this discrimination. For instance, variable importance on projection (VIP) scores can be used for this purpose [34]. VIP scores measure the importance of each predictor variable into the final PLS model and are calculated as the weighted sum of squared PLS variable weights. The “greater than one” rule is used as the criterion to identify the most important variables for a given model because the average of squared VIP scores is equal to 1 [35].

Data were autoscaled prior the application of PLS-DA. This method was applied using PLS Toolbox 8.0.2 (Eigenvector Research Inc, Wenatchee, WA, USA).

### 2.4.5 Lipid identification

Finally, the selected VIP lipids were identified using their MS/MS spectra. These MS/MS spectra were obtained reinjecting rice samples and using as parent ion the most intense *m/z* signal for each previously detected component. These *m/z* signals were retrieved from the MCR-ALS resolved mass spectra corresponding to the VIP selected components.

In untargeted studies, there is no information about what daughter ions should be selected for the analysis. Therefore, optimization of CE was not possible and all MS/MS spectra were recorded at four different CE (10, 20, 30 and 40 eV).

Finally, experimentally obtained MS/MS spectra were compared with *in-silico* MS/MS spectra available in public databases, such as Metlin [36]. The identification was considered satisfactory when at least two daughter ions of the experimental MS/MS spectra coincided with their *m/z* values and their relative abundances with the fragments of *in-silico* MS/MS spectra.

# **3. Results and Discussion**

## 3.1. Resolution of LC×LC-MS lipidomic data

From the large number of m/z values acquired by the MS instrument, ROI strategy selected a relatively low number of *m/z* values for statistical analysis. After applying ROI augmentation, four LC×LC-MS augmented data matrices were obtained as described before: **DAP**, **DAN**, **DRP ,** and **DRN**. Each one of these four column-wise augmented data matrices contained the LC×LC-MS analysis of the studied samples. These matrices had a final size of 63060 rows (total number of retention time channels) and 1505, 1502, 1482 and 1225 columns (number of finally selected *m/z* ROI values), respectively.

Figure 3 shows an example of an obtained LC×LC-MS ROI chromatogram for the control aerial part of one of the rice samples analyzed in positive mode. In this Figure, the chromatograms obtained for each ROI *m/z* values are depicted in different colors. This figure reveals that despite the data compression achieved by selection of mass traces by the ROI procedure, LC×LC-MS chromatograms still presented rather complex profiles with multiple coeluted compounds. Consequently, the detection and identification of lipids cannot be performed directly. For this reason, the application of the MCR-ALS procedure was proposed to get a deeper insight into this experimental lipidomic data.

Quality assessment of the chromatographic runs was performed using QC samples (see Methods section). Instrumental conditions during the entire LC×LC-MS sequence were stable, and therefore, no further corrections were required.

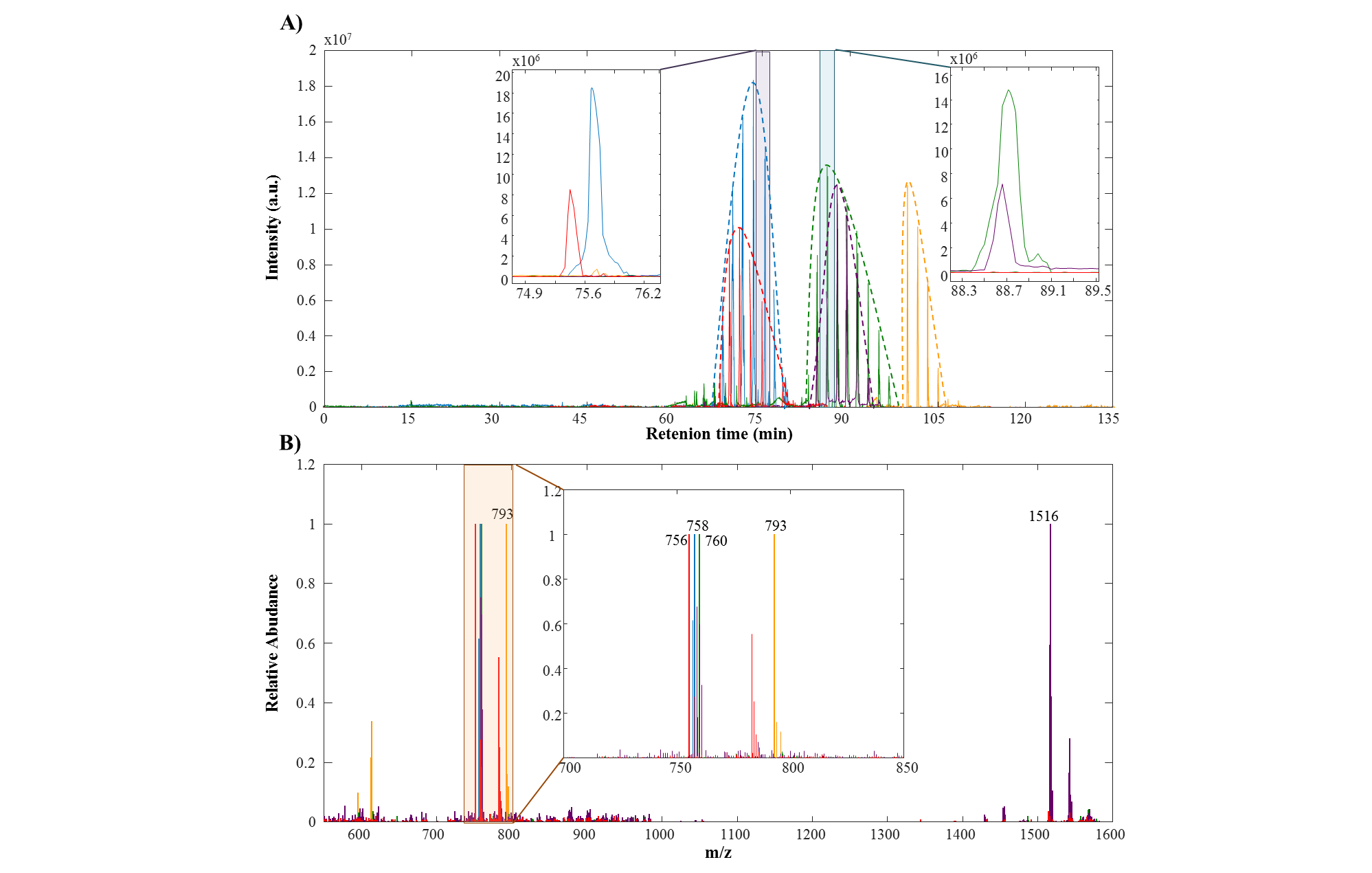
**Figure 3.** Example of a processed LC×LC-MS ROI chromatogram: control aerial part sample analyzed in positive mode. Different colors show the chromatograms obtained for each ROI *m/z* value.

Moreover the repeatability of the performed LCLC-MS analysis was assessed using the retention times and the chromatographic areas of the eight extraction standards in all the samples. A relative standard deviation (RSD %) between 1% and 4% was obtained for retention times, and between 8% and 16% for chromatographic areas. Taking into account the high complexity of LCxLC performance the obtained repeatability is acceptable. Supplementary tables S1 and S2 contain the values of retention times and chromatographic areas of the standards in all the analyzed samples.

In order to resolve the pure elution profiles of the lipids present in the analyzed sample in both chromatographic dimensions and their mass spectra, MCR-ALS was applied separately to the four data matrices obtained in the analysis of the aerial parts and roots of the rice samples, both in positive and negative ionization modes (**DAP**, **DAN**, **DRP** and **DRN**). These four matrices were resolved by an MCR-ALS model using a total number of 250 components. This large number of resolved components included all detected lipid contributions as well as other noisy chromatographic signals, such as instrumental background and solvent contributions. In this case, approximately 200 resolved components were finally assigned to individual lipids for each one of the analyzed matrices. The percentage of explained variance (R2) and the lack of fit (LOF) were considered satisfactory. For instance, matrix **DAP** was resolved with a total variance explained (R2) of 97.3% and with a lack of fit (LOF) equal to 16.5%. Similar results were obtained for the other three matrices, which are given in Supplementary Information.

Figure 4 shows an example of MCR-ALS resolution of the pure elution and mass spectra profiles of five lipids in the aerial part of the rice samples analyzed in positive mode (**DAP** matrix). Figure 4A shows the resolved elution profiles. The dashed lines represent the elution profiles in the first chromatographic dimension, which can be recovered by properly refolding the modulations in the second-dimension column. Each first-dimension peak was divided at least in four modulations, which are represented in Figure 4A with the solid lines, and show the separation achieved at the second-dimension column. Inserts in Figure 4A depict zoomed views of the second-column modulations. Finally, Figure 4B displays the mass spectra of each component.

In Figure 4, the three typical LC×LC-MS elution cases can be clearly appreciated: i) no coelution (yellow signals in Figure 4A); ii) coelution only in the first chromatographic dimension (red and blue signals in Figure 4A); iii) total coelution with embedded peaks (purple and green signals in Figure 4A). These three situations are detailed below.

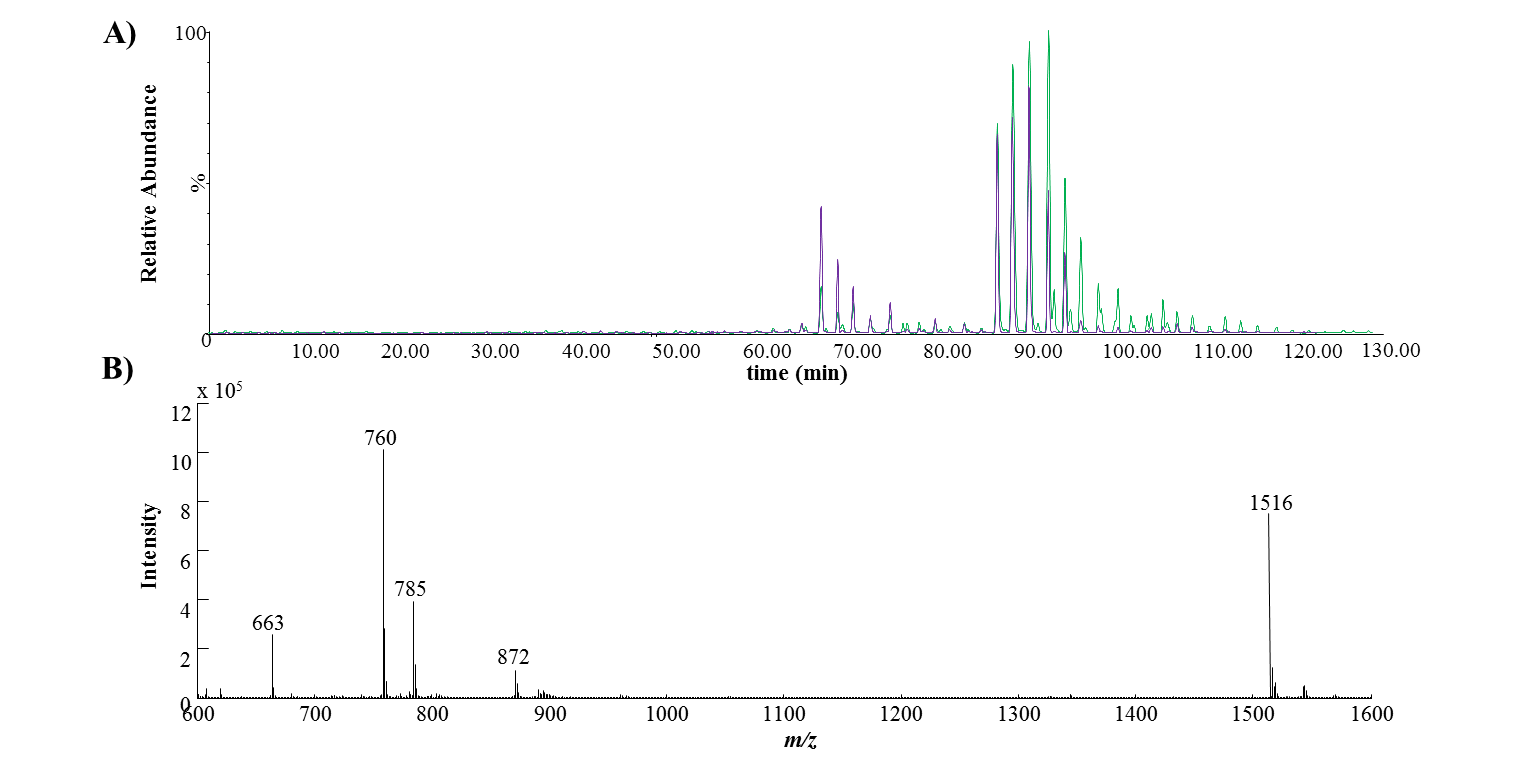


**Figure 4.** Example of MCR-ALS results of five lipids in aerial part samples analyzed in positive mode. A) Pure resolved elution profiles. Inserts show zoomed views of elution profiles between 74.9 and 76.2 minutes and between 93.3 and 94.1 minutes. B) Pure resolved mass spectra. Insert shows zoomed view of mass spectra between 700 and 850 *m/z*. Yellow signals correspond to a non coeluted lipid. Red and blue signals show an example of two lipipids coeluting only in the first chromatographic dimension. Purple and green signals are two totally coeluted lipids with embedded peaks. The no coelution case was the simplest one and could be easily resolved by traditional means. Signals colored in yellow in Figure 4A are an example of this case of no coelution. The four yellow chromatographic peaks are the four modulations in the second-dimension column of a lipid that eluted between 100 and 105 minutes of the first-dimension column. The resolved mass spectrum of this lipid is colored in yellow in Figure 4B. This spectrum allowed the selection of the ion at 793 *m/z* as the parent ion to be used for MS/MS analysis.

The two chromatographic peaks eluted at 75 minutes (blue and red signals in Figure 4A) were an example of coelution only in the first chromatographic dimension (dashed lines). Figure 4A shows that these two lipids coeluted in the first chromatographic column, both eluted between 70 and 80 minutes. However, as shown in the zoomed view, in the second chromatographic dimension the two lipids were separated. This example revealed one benefit of using LC×LC instead of LC, because their chromatographic separation was achieved on the second dimension. Resolved mass spectra of these lipids are colored in blue and red in Figure 4B. In this case, the selected parent ions were 756 and 758 *m/z*.

Finally, the two chromatographic peaks present at 90 minutes (purple and green signals in Figure 4A) were an example of the total coelution case. In Figure 4A it can be seen that these two lipids eluted between 85 and 95 minutes from the first chromatographic column. Moreover, the zoomed view shows that they were also coeluted in the second chromatographic column. This example exposed the large complexity of samples analyzed in lipidomic studies. Despite the high separation power of LC×LC, a total separation of all the lipids present in rice samples could not be achieved. In these cases, the use of chemometric tools such as MCR-ALS is mandatory. As shown in Figure 4, MCR-ALS was capable of resolving the pure elution and mass spectra profiles of these totally coeluted lipids. The resolved mass spectra of these lipids are colored in green and purple in Figure 4B. In this last case, the selected parent ions were 760 and 1516 *m/z*.

Figure 5 shows the comparison of the results obtained using the ROIMCR approach and by manual inspection of the detected signals considering the hardest case of total coelution. Figure 5A shows the raw extracted chromatograms for *m/z* values 760 (colored in green) and 1516 (colored in purple). In this figure, the chromatographic modulations for both *m/z* values appeared at two different retentions times, one at 70 minutes and the other at 90 minutes. On the contrary, the MCR-ALS resolved profiles for these mass traces (green and purple profiles in Figure 4A) only showed chromatographic modulations at 90 minutes. This result indicated that probably at 70 minutes eluted other lipids that have 760 and 1516 *m/z* values as minor signals on their mass spectra. One of the benefits of using MCR-ALS was that it gives the chromatographic signals for each resolved component (lipids). Moreover, Figure 5B shows the raw mass spectrum obtained between 85 and 95 minutes, which contains five intense mass signals: 663, 760, 785, 872 and 1516 *m/z*. This MS spectrum demonstrated that when the manual inspection is used, it was difficult to determine which of these mass traces belong to the same lipid. On the contrary, when ROIMCR approach was used, it gave the pure mass spectra for each of the resolved lipids. In addition, it should be highlighted that a manual inspection of the detected signals is extremely time-consuming because all intense mass traces detected at every retention time should be checked individually. On the contrary, ROIMCR approach allowed a rapid resolution of all of them in the entire dataset.

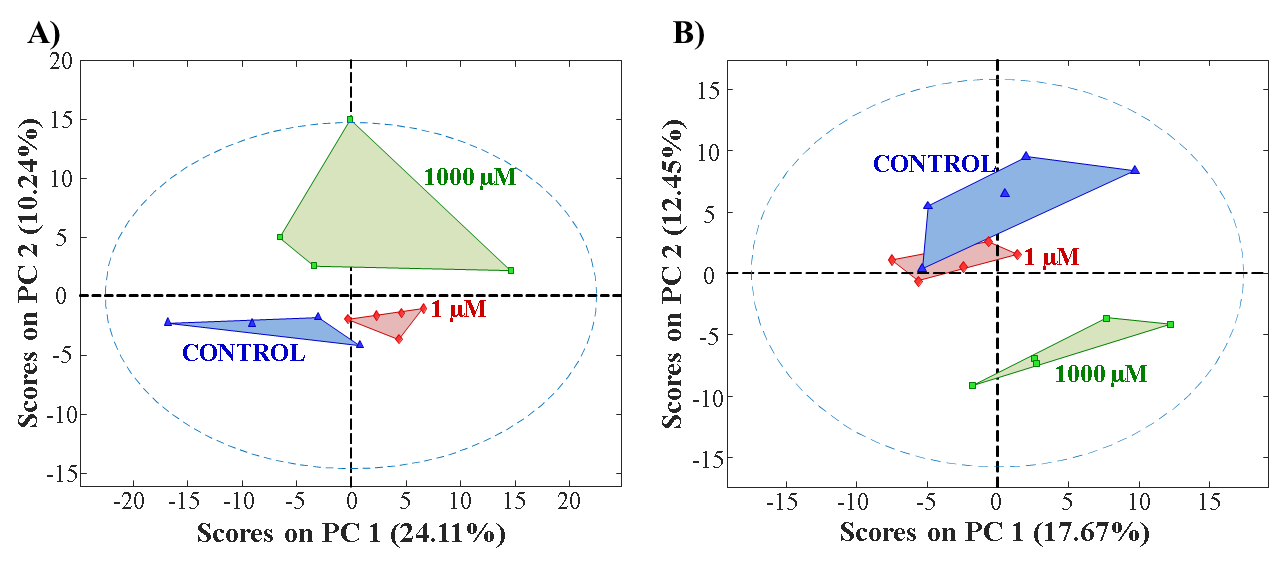


**Figure 5.** A) Raw extracted chromatograms for *m/z* values 760 (colored in green) and 1516 (colored in purple). B) Raw mass spectrum obtained between 85 and 95 minutes.

## 3.2. Statistical assessment of As effects on rice lipidome

The application of one-way ASCA model to every peak area data matrix (**AAP**, **AAN**, **ARP**, and **ARN**) revealed that in the four cases the effects produced by As exposure were significant with *p*-values between 0.003 and 0.001.

The application of PCA in the four cases also showed the effects of As exposure. In all cases, control samples were distinguished from samples exposed to 1000 µM As treatment using first and second principal components (PCs), which already explained more than the 30% of all data variance. For instance, Figure 6 shows PCA scores plots for aerial part (Figure 6A) and root (Figure 6B) samples analyzed in positive ionization mode. In the case of aerial part samples (Figure 6A) PC1 slightly separated control samples from 1µM As watered samples. In contrast, PC2 separates samples watered at 1000 µM. PCA scores plot for root samples (Figure 6B) differentiated samples exposed to 1000 µM As from the others also along PC2. However, in this case, control samples were not distinguished from samples exposed to 1 µM As exposure. These results indicated that As exposure at high doses (1000 µM) affected the rice lipidome, but these effects were low when considering a concentration under the limit accepted by European legislation (1 µM) [22].



**Figure 6.** PCA scores plot obtained for A) aerial part samples and B) root samples analyzed in positive mode.

Results obtained for samples analyzed in negative mode are shown in Supplementary Information (Figure S3). From these results, it can be mentioned that samples treated with the lower concentration of As (1 µM) were only clearly distinguished from control samples in the PCA model of roots analyzed in negative mode (Figure S3B).

## 3.3. Feature Selection

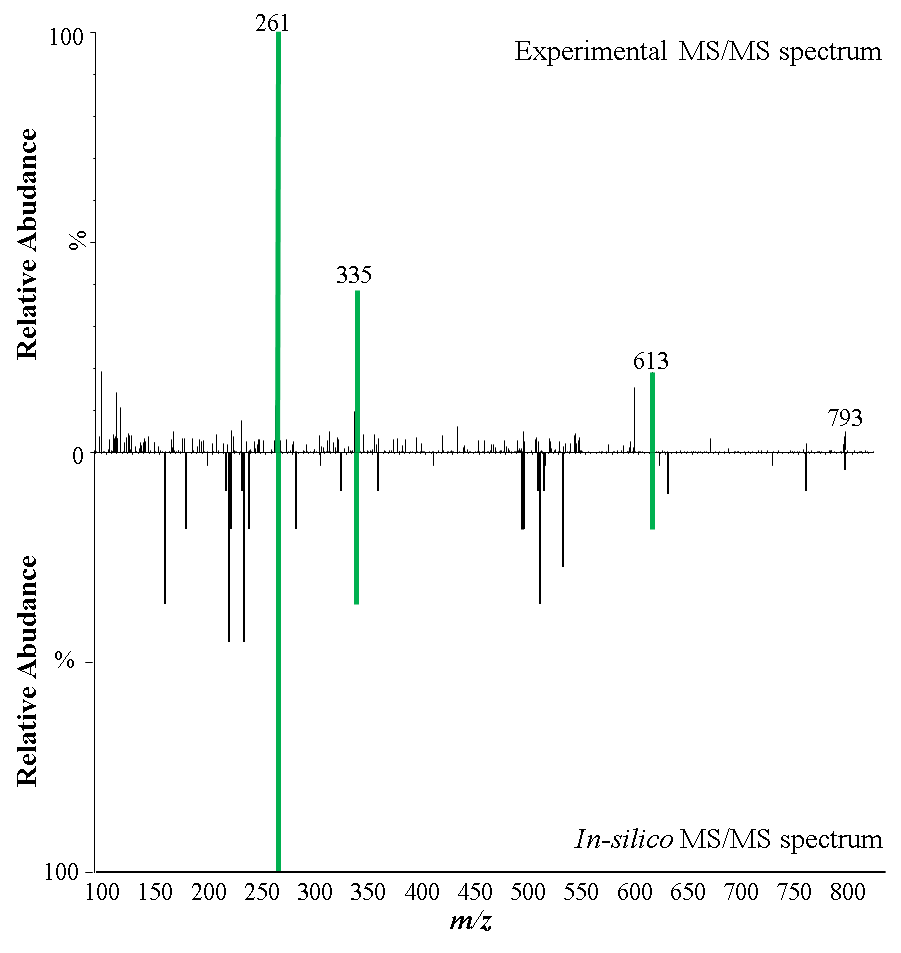
The last step of the chemometric based-strategy for LC×LC-MS data analysis was the selection of the relevant lipids for sample discrimination. These important lipids were those that suffered a significant change under As exposure and, therefore, allowed the differentiation between control and treated samples. With this purpose, PLS-DA was applied to the four peak area matrices (**AAP**,**AAN**,**ARP,** and**ARN**), but only taking into account control and high dose As (1000 µM) treated samples.

The four obtained PLS-DA models distinguished control from treated samples. As an example, Figure S4 in Supplementary Information shows the PLS-DA results for root samples analyzed in negative mode. Figure S4A represents the cross-validation (CV) class predictions and shows that all samples from both classes (control and treated) were perfectly discriminated. Figure S4B shows the VIP scores plot for the mentioned PLS-DA model. The variables (resolved lipids) with a VIP value greater than one were considered important to discriminate among As exposure factor levels. In the case of the example in Figure S4B (roots analyzed in negative mode), a total of 74 lipids were selected. For roots analyzed in the positive mode the number of selected variables was 54. Finally, in the case of aerial part samples, the number of selected lipids was 70 and 77 for samples analyzed in positive and negative modes, respectively.

## 3.3. Lipids identification

In order to identify the lipids whose concentration changed under As exposure, samples were reanalyzed by MS/MS. As mentioned above, lipids were identified by comparison of their experimental MS/MS spectra (recorded at 10, 20, 30 and 40 eV CE) with *in-silico* theoretical spectra available in the METLIN database [36].

Figure 7 shows an example of this MS/MS identification. The upper part of this figure represents the experimental MS/MS spectra at 20 eV CE of the lipid eluted at 100 minutes in Figure 4A (yellow signals in Figure 4). The *m/z* value of the parent ion obtained from the MCR-ALS resolved mass spectrum was 793 (see Figure 4B). This experimental MS/MS spectrum showed three major product ions at 261, 335 and 613 *m/z*, which were well correlated with the theoretical product ions of MGDG (36:6) (METLIN ID 75584).



**Figure 7.** Identification of MCR-ALS resolved component number 5 for aerial part samples analyzed in positive mode as MGDG (36:6). The *m/z* value of the parent ion was 793. The experimental MS/MS spectra for this ion obtained at 20 eV CE (up signals) could be associated with the *in-silico* mass spectra of MGDG (36:6) obtained from Metlin database (down signals, METLIN ID 75584). The major product ions signals are colored in green.

Since the main goal of this work was not to perform a detailed biological interpretation of the observed lipid changes, only the identification of some of the lipids encountered was atempted to confirm the reliability of the proposed methodology. Results of this identification are shown in Table 1, with the *m/z* values of parent ions, the name of the identified lipids, their elemental composition and METLIN ID, the product ions that allowed their identification and the CE (collision energy) of the MS/MS spectrum used for the identification.

**Table 1.** Summary of results obtained for lipid identification.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Parent Ion** | **Lipid name** | **Elemental composition** | **METLIN ID** | **Product Ions** | **CE (eV)** |
| AERIAL PARTS ESI (+) | 587 | DAG (34:5) | C37H62O5 | 58746 | 209/231 | 40 |
| 756 | PC(34:3) | C42H78NO8P | 39422 | 184/237 | 20 |
| 758 | PC(P-34:3) | C42H78NO7P | 62042 | 184/223 | 20 |
| 760 | PC(P-34:2) | C42H80NO7P | 59605 | 184/263 | 20 |
| 786 | PC(P-36:3) | C44H82NO7P | 59607 | 184/263 | 20 |
| 793 | MGDG(36:6) | C45H74O10 | 75584 | 261/335/613 | 20 |
| 796 | TAG(48:6) | C51H86O6 | 98532 | 261/35 | 20 |
| 815 | PI(O-32:0) | C41H81O12P | 81067 | 181/225 | 20 |
| 873 | TAG(52:4) | C55H98O6 | 4838 | 239/593 | 10 |
| 875 | PI(36:5) | C45H77O13P | 80427 | 575/839 | 20 |
| AERIAL PARTS ESI (-) | 745 | PA(36:2) | C39H73O8P | 81923 | 375/470 | 20 |
| 748 | PG(34:1) | C40H77O10P | 61862 | 265/493 | 20 |
| 820 | MGDG(36:6) | C45H74O10 | 75584 | 495/755 | 10 |
| 837 | PI(P-34:4) | C43H77O12P | 80995 | 95/229 | 40 |
| 841 | PC(P-34:3) | C42H78NO7P | 62042 | 182/259 | 20 |
| ROOTS ESI (+) | 464 | Linolenyl laurate | C30H54O2 | 97081 | 183/247/447 | 20 |
| 573 | DAG(P-32:1) | C35H66O4 | 4697 | 113/265 | 20 |
| 613 | DAG(34:3) | C37H66O5 | 4319 | 237/263 | 20 |
| 778 | PC(36:6) | C44H76NO8P | 39691 | 184/579 | 40 |
| 781 | PC(P-36:4) | C44H80NO7P | 59640 | 184/567 | 40 |
| 783 | PC(36:4) | C44H80NO8P | 59649 | 184/319 | 20 |
| ROOTS ESI (-) | 743 | PG(34:3) | C40H73O10P | 61852 | 237/259 | 20 |
| 804 | PC(34:1) | C42H82NO8P | 39326 | 238/758 | 10 |
| 815 | PC(P-36:2) | C44H84NO7P | 59543 | 263/768 | 10 |
| 821 | PC(36:3) | C44H82NO8P | 39511 | 217/231 | 40 |
| 962 | Glycerol 2-(9Z,12Z-octadecadienoate) 1-hexadecanoate 3-O-[alpha-D-galactopyranosyl-(1->6)-beta-D-galactopyranoside] | C49H88O15 | 95895 | 635/897 | 20 |

As an example of the possibilities of the presented methodology, four of the five lipids shown in Figure 4 were finally identified. The lipid eluted at 100 minutes was identified in Figure 7 as MGDG (36:6). The two lipids coleuting in the first-dimension column, with parent ions at 756 (red signal) and 758 (blue signal) *m/z* values were identified as PC(34:3) and PC(P-34:3), respectively. The MS/MS spectra of the first one showed two product ions at 184 and 237 *m/z* that could be correlated to the theoretical spectra. In the second case, the two product ions used for identifying the lipids were at 184 and 223 *m/z*. Finally, only one of the two totally coeluted lipids could be completely identified. The MS/MS spectrum of the one with the parent ion at 760 *m/z* (green signal) gave two major product ions at 184 and 263 *m/z*, which could be associated with the theoretical product ions of PC(P-34:2). The other one (purple signals in Figure 4), could be tentatively identified as PC(34:2) in agreement with the mass of the parent ion and its retention time. The mass of the parent ion was 1516 Da could be assigned to the [2M+H]+ adduct of PC(34:2). This compound probably coelutes with PC (P-34:2) (green signals in Figure 4A). Unfortunately, the obtained MS/MS spectra for this parent ion could not confirm this identification.

# **Conclusions**

A chemometrics-based data analysis strategy is proposed to gather all relevant information from untargeted lipidomic LC×LC-MS datasets. Despite the high complexity of untargeted LC×LC-MS datasets, the ROIMCR compression and resolution strategy allowed the determination of a large number of lipids and of the changes in their concentration in one single analysis of the LC×LC-MS data from rice samples exposed to As.

The main advantage of the proposed methodology is that it achieves a satisfactory resolution of complex lipidomics samples. In comparison with LC-MS based systems and with LC×LC-MS analysis followed by manual inspection of the detected signals, a higher number of lipids could be resolved. However, the difficulty in the identification of lipids is a relevant drawback. On the one hand, there is still a lack of theoretical lipid MS/MS spectra available in public databases. On the other hand, some of the obtained MS/MS spectra did not show daughter ions withintense enough *m/z* signals. This may be related to a poor detection sensitivity resulting from the dilution caused by two successive chromatographic steps or to the impossibility of optimizing CE (collision energy). Potential solutions to overcome the detection-sensitivity limitations should be considered. For instance, the use of an active modulation as indicated in the recent work by Gargano [37]. Moreover, after the untargeted detection of the potential lipid biomarkers, the confirmation of these candidates may be performed using a targeted approach with better sensitivity and, consequently, provide a higher number of product ions after, for instance, the optimization of CE. Also, the combination with high-resolution mass spectrometry (HRMS) would improve the detection-sensitivity and reduce the number of candidates to identify allowing a prior tentative identification by exact mass.

Considering the rice lipidomic study performed in this work, results showed that As exposure had significant effects on rice lipidome (specially at the high dose). However, the irrigation of rice plants with water that contains As at a concentration accepted by the European legislation did not show significant effects on rice lipidome.

# **Acknowledgments**

The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement Number 320737.

# **References**

[1] S.M. Lam, G. Shui, Lipidomics as a Principal Tool for Advancing Biomedical Research, Journal of Genetics and Genomics, 40 (2013) 375-390.

[2] S. Sethi, E. Brietzke, Recent advances in lipidomics: Analytical and clinical perspectives, Prostaglandins & Other Lipid Mediators, 128-129 (2017) 8-16.

[3] A. Checa, C. Bedia, J. Jaumot, Lipidomic data analysis: Tutorial, practical guidelines and applications, Analytica Chimica Acta, 885 (2015) 1-16.

[4] F. Fenaille, P. Barbier Saint-Hilaire, K. Rousseau, C. Junot, Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand?, Journal of Chromatography A, 1526 (2017) 1-12.

[5] M. Li, Z. Zhou, H. Nie, Y. Bai, H. Liu, Recent advances of chromatography and mass spectrometry in lipidomics, Analytical and Bioanalytical Chemistry, 399 (2011) 243-249.

[6] C. Wolf, P.J. Quinn, Lipidomics: practical aspects and applications, Progress in lipid research, 47 (2008) 15-36.

[7] N. Navas-Iglesias, A. Carrasco-Pancorbo, L. Cuadros-Rodríguez, From lipids analysis towards lipidomics, a new challenge for the analytical chemistry of the 21st century. Part II: Analytical lipidomics, TrAC Trends in Analytical Chemistry, 28 (2009) 393-403.

[8] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry, TrAC Trends in Analytical Chemistry, 61 (2014) 192-206.

[9] N. Vinayavekhin, A. Saghatelian, Untargeted metabolomics, Current protocols in molecular biology, Chapter 30 (2010) Unit 30.31.31-24.

[10] E. Fahy, S. Subramaniam, R.C. Murphy, M. Nishijima, C.R.H. Raetz, T. Shimizu, F. Spener, G. van Meer, M.J.O. Wakelam, E.A. Dennis, Update of the LIPID MAPS comprehensive classification system for lipids, Journal of Lipid Research, 50 (2009) S9-S14.

[11] A. Baglai, A.F.G. Gargano, J. Jordens, Y. Mengerink, M. Honing, S. van der Wal, P.J. Schoenmakers, Comprehensive lipidomic analysis of human plasma using multidimensional liquid- and gas-phase separations: Two-dimensional liquid chromatography–mass spectrometry vs. liquid chromatography–trapped-ion-mobility–mass spectrometry, Journal of Chromatography A, 1530 (2017) 90-103.

[12] D.R. Stoll, P.W. Carr, Two-Dimensional Liquid Chromatography: A State of the Art Tutorial, Analytical Chemistry, 89 (2017) 519-531.

[13] P. Dugo, F. Cacciola, T. Kumm, G. Dugo, L. Mondello, Comprehensive multidimensional liquid chromatography: Theory and applications, Journal of Chromatography A, 1184 (2008) 353-368.

[14] P.W. Carr, J.M. Davis, S.C. Rutan, D.R. Stoll, Principles of Online Comprehensive Multidimensional Liquid Chromatography, Advances in chromatography, 50 (2012) 139-235.

[15] S.E. Porter, D.R. Stoll, S.C. Rutan, P.W. Carr, J.D. Cohen, Analysis of four-way two-dimensional liquid chromatography-diode array data: application to metabolomics, Anal Chem, 78 (2006) 5559-5569.

[16] M. Navarro-Reig, J. Jaumot, A. Baglai, G. Vivó-Truyols, P.J. Schoenmakers, R. Tauler, Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution, Analytical Chemistry, 89 (2017) 7675-7683.

[17] M.M. Sinanian, D.W. Cook, S.C. Rutan, D.S. Wijesinghe, Multivariate Curve Resolution-Alternating Least Squares Analysis of High-Resolution Liquid Chromatography–Mass Spectrometry Data, Anal. Chem., 88 (2016) 11092-11099.

[18] C. Tistaert, H.P. Bailey, R.C. Allen, Y. Vander Heyden, S.C. Rutan, Resolution of spectrally rank-deficient multivariate curve resolution: Alternating least squares components in comprehensive two-dimensional liquid chromatographic analysis, J. Chemom., 26 (2012) 474-486.

[19] D.W. Cook, S.C. Rutan, D.R. Stoll, P.W. Carr, Two dimensional assisted liquid chromatography - a chemometric approach to improve accuracy and precision of quantitation in liquid chromatography using 2D separation, dual detectors, and multivariate curve resolution, Anal. Chim. Acta., 859 (2015) 87-95.

[20] V. Matyash, G. Liebisch, T.V. Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-terf-butyl ether for high-throughput lipidomics, Journal of Lipid Research, 49 (2008) 1137-1146.

[21] M. Navarro-Reig, J. Jaumot, B. Piña, E. Moyano, M.T. Galceran, R. Tauler, Metabolomic analysis of the effects of cadmium and copper treatment in: Oryza sativa L. using untargeted liquid chromatography coupled to high resolution mass spectrometry and all-ion fragmentation, Metallomics, 9 (2017) 660-675.

[22] D. 2006/118/EC, Directive 2006/118/EC of the European parliment and of the council of 12 December 2006 on the protection of groundwater against pollution and deterioration, in, 2006.

[23] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, TrAC - Trends in Analytical Chemistry, 82 (2016) 425-442.

[24] A. De Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Chim. Acta, 6 (2014) 4964-4976.

[25] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemometrics Intell. Lab. Syst., 140 (2015) 1-12.

[26] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics Intell. Lab. Syst., 2 (1987) 37-52.

[27] J.J. Jansen, H.C.J. Hoefsloot, J. Van Der Greef, M.E. Timmerman, J.A. Westerhuis, A.K. Smilde, ASCA: Analysis of multivariate data obtained from an experimental design, J. Chemometr., 19 (2005) 469-481.

[28] M. Barker, W. Rayens, Partial least squares for discrimination, J. Chemometr., 17 (2003) 166-173.

[29] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: A review of advanced and tailored applications and challenges, Anal. Chim. Acta, 765 (2013) 28-36.

[30] D.J. Vis, J.A. Westerhuis, A.K. Smilde, J. van der Greef, Statistical validation of megavariate effects in ASCA, BMC Bioinformatics, 8 (2007) 322.

[31] A.K. Smilde, J.J. Jansen, H.C.J. Hoefsloot, R.J.A.N. Lamers, J. van der Greef, M.E. Timmerman, ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data, Bioinformatics, 21 (2005) 3043-3048.

[32] D.J. Vis, J.A. Westerhuis, A.K. Smilde, J. van der Greef, Statistical validation of megavariate effects in ASCA, BMC Bioinformatics, 8 (2007).

[33] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta, 185 (1986) 1-17.

[34] S. Wold, A. Johansson, M. Cocchi, PLS-partial least squares projections to latent structures, in: H. Kubiny (Ed.) 3D QSAR in Drug Design, ESCOM Science Publishers, Leiden, 1993, pp. 583-618.

[35] I.G. Chong, C.H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemometrics Intell. Lab. Syst., 78 (2005) 103-112.

[36] R. Tautenhahn, K. Cho, W. Uritboonthai, Z. Zhu, G.J. Patti, G. Siuzdak, An accelerated workflow for untargeted metabolomics using the METLIN database, Nature Biotechnology, 30 (2012) 826.

[37] A.F.G. Gargano, M. Duffin, P. Navarro, P.J. Schoenmakers, Reducing Dilution and Analysis Time in Online Comprehensive Two-Dimensional Liquid Chromatography by Active Modulation, Analytical Chemistry, 88 (2016) 1785-1793.