Chemometric analysis of comprehensive LC×LC-MS data: resolution of triacylglycerol structural isomers in corn oil

Meritxell Navarro-Reig, Joaquim Jaumot, Teris A. van Beek, Gabriel Vivó-Truyols, Romà Tauler

www.elsevier.com/locate/talanta

PII: S0039-9140(16)30570-7
DOI: http://dx.doi.org/10.1016/j.talanta.2016.08.005
Reference: TAL16769

To appear in: *Talanta*

Received date: 12 May 2016
Revised date: 28 July 2016
Accepted date: 1 August 2016

Cite this article as: Meritxell Navarro-Reig, Joaquim Jaumot, Teris A. van Beek Gabriel Vivó-Truyols and Romà Tauler, Chemometric analysis of comprehensive LC×LC-MS data: resolution of triacylglycerol structural isomers in corn oil *Talanta,* http://dx.doi.org/10.1016/j.talanta.2016.08.005

Chemometric analysis of comprehensive LC×LC-MS data: resolution of triacylglycerol structural isomers in corn oil

Meritxell Navarro-Reig[a], Joaquim Jaumot[a], Teris A. van Beek[b], Gabriel Vivó-Truyols[c], Romà Tauler[a*]

a) IDAEA-CSIC, Department of Environmental Chemistry, Barcelona, Spain

b) Laboratory of Organic Chemistry, Natural Products Chemistry Group, Wageningen University, Wageningen, The Netherlands

c) Universiteit van Amsterdam, van't Hoff Institute for Molecular Science, Amsterdam, The Netherlands

Corresponding author: Romà Tauler, Professor

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

Telephone: +34934006140

E-mail: Roma.Tauler@idaea.csic.es

Abstract

Comprehensive hyphenated two-dimensional liquid chromatography mass spectrometry (LC×LC-MS) is a very powerful analytical tool achieving high throughput resolution of highly complex natural samples. However, even using this approach there is still the possibility of not resolving some of the analytes of interest. For instance, triacylglycerols (TAGs) structural isomers in oil samples are extremely difficult to separate chromatographically due to their very similar structure and chemical properties. Traditional approaches based on current vendor chromatographic software cannot distinguish these isomers from their different mass spectral features. In this work, a chemometric approach is proposed to solve this problem. First, the experimental LC×LC-MS data structure is discussed, and results achieved by different methods based on the fulfilment of the trilinear model are compared. Then, the step-by-step resolution and identification of strongly coeluted compounds from different examples of triacylglycerols (TAGs) structural isomers in corn oil samples are described. As a conclusion, the separation power of two-dimensional chromatography can be significantly improved when it is combined with the multivariate curve resolution method.

Keywords:

LC×LC-MS, triacylglycerols, structural isomers, trilinearity, PARAFAC, MCR-ALS

1. Introduction

Over the past years, the need for analysis of complex samples has increased considerably in a broad variety of fields, including environmental, clinical or food industry. Liquid chromatography is often the analytical approach chosen for analysing those samples, due to its high ability for resolving complex mixtures. Nevertheless, one-dimension chromatography is not always capable of separating all constituents in complex natural samples. Comprehensive two-dimensional liquid chromatography (LC×LC) appears as a powerful alternative to achieve a better separation of all these constituents [1]. The high resolving power of this technique lies in the fact that under ideal circumstances, when retention mechanisms of the two separation dimensions are uncorrelated, the overall peak capacity is equal to the product of the individual peak capacities of the first and second dimension separations [1, 2]. For this reason, during the last few years, major attention has been focused on the development of comprehensive two-dimensional liquid chromatography methodologies coupled to multivariate detectors. For example, molecular absorption diode array or mass spectrometric detectors have been used to deal with complex samples, such as egg yolk, urine, urban aerosols, red wine or polymers, among others [1-9].

Although comprehensive two-dimensional liquid chromatography provides a better resolution than one-dimensional chromatography, there is still the possibility that some of the analytes in complex samples remain unresolved. One example of these limitations is the analysis of triacylglycerols (TAGs) in vegetable oils. Different combinations of three fatty acids connected to the glycerol backbone generate different TAG structural isomers, which can be positional or chain isomers. Positional isomers are generated by variations in the position of the same three fatty acids on the glycerol molecule, whereas chain isomers refer to the different chain lengths, the number of double bonds or position of double bonds of the different fatty acids. Figure S1 on supplementary material shows examples of these structural isomers, which have similar chromatographic behaviour. Therefore, their chromatographic separation is highly complex. Moreover, the distinction between structural isomers by their mass spectra is also troublesome.

Two TAGs chain isomers can have the same *m/z* value for their protonated molecular ion and some of their diacylglycerols fragments. The case of positional isomers is even more troublesome, because they have the same *m/z* value for their molecular ion and all of their diacylglycerols fragments. Therefore, the achievement of a complete resolution of these compounds is important, especially for those studies which aim at separating multiple compounds with similar structures, such as TAGs in vegetable oils [10].

It is logical to hypothesize that the combination of chemometric methods with comprehensive two-dimensional liquid chromatography could achieve a better resolution of constituents in complex natural samples. There are already multiple examples of the application of chemometric methods to analyse one-dimensional chromatographic data [11-14]. However, in this work, we will focus our study on the analysis of two-dimensional liquid chromatographic data.

Most of the contributions found in the literature dealing with curve resolution of multidimensional separations consider two-dimensional gas chromatography (GC×GC). In this field, it is noteworthy the work done by Synovec [15-18] and by Parastar [19-23]. In contrast to two-dimensional gas chromatography, comprehensive two-dimensional liquid chromatography data analysis is principally done manually using vendor software tools, and the use of chemometric tools for peak detection and resolution is a relatively new concept in this field. For this reason, there are only a few references of using chemometric tools to analyse LC×LC data. In this area, the work of Rutan and co-workers should be highlighted. They have developed a methodology to analyse LC×LC-DAD data based on the application of the multivariate curve resolution–alternating least squares (MCR-ALS) method [24-26]. Other attempts at designing algorithms for detection and quantification of peaks can also be mentioned [27, 28]. To the best of our knowledge, the development of a general strategy for the chemometric analysis of LC×LC-MS data for chromatographic peak detection and quantification is still a challenge.

From a chemometric point of view, LC×LC-MS provides three-way data cube and the curve resolution based on trilinear models, such as parallel factor analysis (PARAFAC), might seem appropriate to resolve this type of data. For instance, Synovec has extensively studied the application of the parallel factor analysis (PARAFAC) method to the analysis of GC×GC-MS [15-18]. However, frequent changes in chromatographic peak shapes and in column retention time shifts between consecutive modulations cause the failure in the fulfilment of the trilinear model postulated in these studies. For this reason, bilinear chemometric methods, such as MCR-ALS, emerge as a valid alternative. As stated above, Rutan [9, 24-26] and Parastar [19-23] have demonstrated the effectiveness of the MCR-ALS method to deal with LC×LC-DAD and GC×GC-MS data, respectively. Recently, Bortolato and Olivieri [29] compared the application of PARAFAC2 and MCR-ALS for the analysis of chromatographic data considering the effects of changes in peak shapes and shifts. The authors concluded that, in cases where strong coelutions and interferences are present (i.e. in natural samples), successful results could only be obtained using MCR-ALS and they also exposed the limitations of the PARAFAC2 method.

In the presented work, two main goals are pursued. First, the development of a strategy to analyse LC×LC-MS based on the application of curve resolution methods taking as a case of study the complete resolution of TAG structural isomers from vegetable oil samples. Additionally, the study of LC×LC-MS data structure is presented, and consequently the possible use of bilinear and trilinear model based methods for the analysis of LC×LC-MS data is evaluated. Results achieved by bilinear methods are compared with those obtained by methods based on trilinear models.

2. Materials and Methods

2.1. LC×LC-MS of triacylglycerols (TAGs) in corn oil samples

In this work, triacylglycerols (TAGs) in corn oil samples were analysed by LC×LC-MS. In the first-chromatographic dimension, an Ag(I)-coated cation exchange (250×2.1 mm, 5 μm) column was used, and in the second chromatographic dimension, a C18 (30x4.6 mm, 1.8 μm) was

employed. Modulations of one minute sections from the first dimension column were introduced into the second dimension column by means of a 10-way valve with two switching loops. Detection was performed by an atmospheric pressure chemical ionization mass spectrometer (APCI-MS) in positive ion mode, and the selected mass spectral range was from *m/z* 250 to 1150 at 1 unit mass resolution.

Figure 1 near here

Figure 1 describes two possible LC×LC-MS data arrangements obtained in the analysis of a single sample. In the first option, shown in Figure 1A, the experimental data is arranged in a so-called third-order tensor, i.e. a data cube, **D**. Full scan mass spectra are in the *x*-axis direction of the same data cube, with 901 channels (from 250 to 1150 *m/z* at 1 unit mass resolution). The first-chromatographic dimension is on the *z*-axis direction (from 0 to 179 min) and the second chromatographic dimension is on the *y*-axis direction (from 0 to 60 seconds). A preliminary data exploration can be done considering different 2D slices of the data cube separately. For instance, a slice of the cube at a particular *m/z* value gives a 2D chromatogram displaying the compounds present at this *m/z* value. A 2D TIC (total ion current) chromatogram can be obtained when all *m/z* intensity values for both chromatographic dimensions are summed. This 2D TIC plot displays the major data features and provides information about retention time regions of more important chromatographic peaks. It is also possible to study individual mass spectra for any combination of the two chromatographic dimensions. Finally, slices at a particular first-dimension elution time show the different chromatographic modulations. Every modulation gives a second dimension chromatographic separation. Therefore, a data matrix can be built for each modulation ($\mathbf{D_K}$ matrix in Figure 1, where *K* is the number of modulations). These $\mathbf{D_K}$ data matrices are the second dimension LC-MS chromatograms at the different *m/z* values. The rows of this data matrix have the mass spectra at every second dimension retention time, and the columns of this data matrix have the second dimension chromatograms at each *m/z* channel.

Figure 1B shows a different LC×LC-MS data arrangement in a column-wise augmented data matrix structure. This augmented data matrix ($\mathbf{D_{aug}}$ matrix in Figure 1B) can be built settling individual $\mathbf{D_K}$ matrices from each modulation one on the top of the other, and keeping *m/z* mode (MS spectra) in common. $\mathbf{D_{aug}}$ matrix contains 901 *m/z* values on its columns (from 250 to 1150 *m/z* at 1 resolution) and 26492 retention times on its rows, which results from the 179 modulations taken from the first-chromatographic dimension multiplied by the 148 second dimension column retention times in the second chromatographic dimension.

2.2. Data preparation

Acquisition times in the second chromatographic dimension were not an exact multiple of the modulation period, and, therefore, the number of measured mass spectra in each modulation was not constant. In order to construct the data cube **D**, it is necessary that the same number of mass spectra is measured in every modulation. Therefore, the optimal measurement period in the second dimension column was calculated as follows:

$$m = t_m / mean(p) \quad \text{Eq. (1)}$$

$$m_2 = round(m) \quad \text{Eq. (2)}$$

$$p_2 = t_m / m_2 \quad \text{Eq. (3)}$$

Where *m* is the mean of the number of acquisition times in the second dimension column within each modulation, t_m is the modulation time, mean(p) is the mean of acquisition times in the second dimension column, m_2 is the rounded number of acquisition times within each modulation, and p_2 is the new period (seconds between two consecutive scans) considered. Then, using this new constant period a new acquisition time axis for the second dimension column can be obtained. This axis will be now constant for all modulations. The last step was then the application of a normal kernel smoothing filter with a bandwidth equal to 0.5 to interpolate the signal intensity of every *m/z* in these newly considered time intervals [30].

2.3. Chemometrics methodology

Two families of chemometric methods have been proposed for the analysis of 2D LC×LC-MS data sets. On the one hand, there are methods that assume that 2D chromatographic data follows a trilinear model such as PARAFAC [31, 32] and PARAFAC2 [33, 34] methods. On the other hand, there are the chemometric methods based on the basic assumption that 2D chromatographic data fulfils the bilinear model (and not necessarily the trilinear model), such as MCR-ALS [35]. The fulfilment of trilinearity in the case of multidimensional chromatographic (i.e. GC×GC or LC×LC) data is at present a controversial topic, and both types of chemometric methods have been previously used to deal with GC×GC [18, 19] and LC×LC data [24, 36]. For this reason, in this work, both, bilinear methods (MCR-ALS bilinear), and trilinear methods (PARAFAC, PARAFAC2, MCR-ALS trilinear, MCR-ALS trilinear allowing time shifting) were tested and compared in the analysis of LC×LC-MS data.

2.3.1. MCR-ALS based methods

MCR-ALS is a chemometric method used for the resolution of the contributions of the pure components present in unresolved complex mixtures. MCR-ALS has been used to investigate a wide variety of problems from different fields, in particular multidimensional chromatographic systems [19, 24]. MCR-ALS has been extensively described in the literature [37-40] and it is only briefly explained here for the particular case of LC×LC-MS data resolution, with the goal of the resolution of the pure elution profiles in both chromatographic dimensions and the pure mass spectra profiles of the constituents of the analysed sample.

MCR-ALS decomposes experimental data according to a bilinear additive model defined by the multi-sample and multi-wavelength generalization of Lambert-Beer's law for spectroscopic measurements. For a data matrix, such as the one taken from one modulation of the first dimension column in LC×LC-MS, this bilinear model can be written as:

$$\mathbf{D_K} = \mathbf{C_K S^T} + \mathbf{E_K} \qquad \text{Eq. (4)}$$

Where $\mathbf{D_K}$ ($I \times J$) is the experimental data matrix corresponding to one of the second dimension column modulations taken from the first dimension column (where I is the number of rows corresponding to the number of retention times or to the number of mass spectra measured at these retention times, and J is the number of columns corresponding to the *m/z* values of the mass spectra), see Figure 1B. $\mathbf{C_K}$ (I × N) is the matrix containing the resolved second dimension elution profiles for this modulation (where N is the number of resolved components by MCR-ALS), and $\mathbf{S^T}$ ($N \times J$) has the mass spectra of these N components. Finally, $\mathbf{E_K}$ (I × J) is the matrix of residuals not explained by the MCR model.

This data analysis strategy can be easily extended to the simultaneous analysis of several second dimension column modulations [19]. The same MCR bilinear model described in Eq. 4 can be extended in this case as follows (Fig. 1B):

$$\mathbf{D_{aug}} = \begin{bmatrix} \mathbf{D_1} \\ \mathbf{D_2} \\ \mathbf{D_3} \\ \vdots \\ \mathbf{D_K} \end{bmatrix} = \begin{bmatrix} \mathbf{C_1} \\ \mathbf{C_2} \\ \mathbf{C_3} \\ \vdots \\ \mathbf{C_K} \end{bmatrix} \mathbf{S^T} + \begin{bmatrix} \mathbf{E_1} \\ \mathbf{E_2} \\ \mathbf{E_3} \\ \vdots \\ \mathbf{E_K} \end{bmatrix} = \mathbf{C_{aug}} \mathbf{S^T} + \mathbf{E_{aug}} \qquad \text{Eq. (5)}$$

Where $\mathbf{D_{aug}}$ is the column-wise augmented data matrix containing multiple second dimension modulations. Since K is the number of the second dimension column modulations taken from the first dimension column, the number of rows for $\mathbf{D_{aug}}$ is equal to $I \times K$. Decomposition of matrix $\mathbf{D_{aug}}$ generates $\mathbf{C_{aug}}$ ($IK \times N$) which has second dimension resolved elution profiles at each modulation for the N resolved components. In addition, $\mathbf{S^T}$ ($N \times J$) represents the mass spectra resolved for the N components, common to all the considered modulations, which can be used to identify them. $\mathbf{E_{aug}}$ ($IK \times J$) has the residuals not explained by the model. First dimension elution profiles for every component can be obtained by refolding appropriately every column in $\mathbf{C_{aug}}$ to give a matrix of dimensions ($I \times K$). Every column of the refolded matrix will give the first-dimension elution profile of size (1, K) and, therefore, the matrix of the first-dimension elution profiles of dimensions ($N \times K$) is obtained.

An initial guess of the number of components of $\mathbf{D_{aug}}$ matrix was obtained using singular value decomposition (SVD) algorithm [41]. Initial estimates of pure component spectra ($\mathbf{S^T}$) profiles were computed using a purest spectra detection method based on SIMPLISMA [42, 43]. Finally, ALS optimization was carried out applying non-negativity (elution and spectra profiles), spectral normalization (equal height) and spectral equality constraints [44, 45].

The described bilinear model for 2D multidimensional chromatographic data can be extended to the trilinear model as follows:

$$\mathbf{D_K} = \mathbf{CT_K S^T} + \mathbf{E_K} \qquad \text{Eq. (6)}$$

Where $\mathbf{D_K}$ is the data matrix for the modulation *K*, **C** is the matrix of the elution profiles of the resolved components, $\mathbf{T_K}$ is a diagonal matrix giving the relative compositions of these components in this particular modulation *K*, and $\mathbf{S^T}$ has their spectra profiles. In this trilinear model, all $\mathbf{C_K}$ matrices in Eq. (6) are assumed to follow the equation $\mathbf{C_K} = \mathbf{CT_K}$. This implies that profiles in $\mathbf{C_K}$ have exactly the same shape and only can differ in their intensity, defined by a scalar factor in diagonal matrix $\mathbf{T_K}$. From a chromatographic viewpoint, this trilinear model implies that the resolved components have the same elution profile in all the *K* modulations with exactly the same retention time. MCR-ALS optionally allows imposing the fulfilment of the trilinear model for all or some of the components during the ALS optimization, and also it allows for some between run shift deviations but keeping the same shape. MCR-ALS trilinear forces that all elution profiles of the same component in different modulations have equal shape and appear at the same retention time. MCR-ALS trilinear allowing time shifting, (i.e. partially trilinear), forces also equal shapes in elution profiles but allows variations in retention time among different chromatographic runs.

In all the cases, ALS optimization convergence criterion of the ALS optimization was set to 0.1 (in % of change of standard of deviation of residuals between two consecutive iterations).

2.3.2. PARAFAC based methods

PARAFAC has been extensively used in the resolution of multi-way data and claims the uniqueness of obtained solutions as one of its main advantages. PARAFAC models are based on the assumption that a data cube, **D**, can be decomposed into a trilinear combination of pure component responses in each of the three modes. PARAFAC has been proposed for chromatographic data [15-17, 32], but the trilinear requirement restricts its use to data where no shift in retention times of the elution profiles of the same component occur among runs and no either peak shape changes are produced. However, these two conditions might be too strong in the current chromatographic practice. PARAFAC has been extensively described in the literature [46-48], and also its application to multidimensional chromatographic data [32, 36].

In this work, PARAFAC was initialized using loadings initial estimates obtained by trilinear decomposition (TLD) [49] and the ALS optimization was carried out under non-negativity constraints in the three modes (elution in both chromatographic dimensions and mass spectra profiles) [31].

PARAFAC2 is a variant of PARAFAC method developed by Bro and co-workers [33, 34] to deal with three-way data with small shifts in the profiles of one of the data modes. Regarding LC×LC data, PARAFAC2 can supposedly handle small retention time shifts across modulations by allowing a certain freedom in the second dimension elution profiles in matrix **C** (see PARAFAC model [31]). Therefore, PARAFAC2 provides as many different $\mathbf{C_k}$ matrices as *K* number of modulations. To keep uniqueness in the solution, the cross-product of different $\mathbf{C}_k$ has to be constant over all modulations ($\mathbf{C_1}\,\mathbf{C^T_1} = \mathbf{C_2}\,\mathbf{C^T_2} = \ldots = \mathbf{C_k C^T_k}$). In PARAFAC2 constraints cannot be applied in this second dimension chromatographic mode. Consequently, non-negativity constraints can only be applied to mass spectra and first-dimension elution profiles.

The comparison of the capability of PARAFAC2 and MCR-ALS resolving chromatographic data is currently a topic of discussion. Recently, Bortolato and Olivieri [29] compared both

methods considering the effects of retention time shifts and changes in peak shapes. The authors concluded that PARAFAC2 would only be able to resolve two overlapped peaks if the time shifts of the two peaks are limited, and no significant changes occur in the profile shapes. These conditions are difficult to accomplish in two-dimensional liquid chromatography. Therefore the study of the structure of LC×LC-MS data is important to know if PARAFAC2 will be able to resolve it.

In all the PARAFAC based analysis the convergence criterion used was the relative change in fit, and was set to 10^{-6}.

2.3.3. Evaluation of the quality of chemometric methods

The quality of the applied methods was assessed using the explained data variance (R^2) and the lack of fit (*LOF*) for both bilinear and trilinear methods. The equations defining these two parameters are:

$$R^2{}_{2-\text{way}}(\%) = 100\sqrt{\frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad R^2{}_{3-\text{way}}(\%) = 100\sqrt{\frac{\sum_{i,j,k} d_{ijk}^2 - \sum_{i,j,k} e_{ijk}^2}{\sum_{i,j,k} d_{ijk}^2}} \qquad \text{Eq. (7)}$$

$$LOF_{2-\text{way}}\ (\%) = 100\sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad LOF_{3-\text{way}}\ (\%) = 100\sqrt{\frac{\sum_{i,j,k} e_{ijk}^2}{\sum_{i,j,k} d_{ijk}^2}} \qquad \text{Eq. (8)}$$

Where in the case of two-way arrays $d_{i,j}$ is an element of the experimental data matrix ($\mathbf{D_{aug}}$) and $e_{i,j}$ is the related residual. And, in the case of three-way array d_{ijk} is an element of the experimental data cube ($\mathbf{D}$) and e_{ijk} is the related residual.

In order to compare between the mass spectra profiles resolved by the different tested methods in a pairwise mode, angle values between the obtained spectra profiles were calculated. Angle value (α) can be calculated taking into account the cosine of the angle between the two vectors representing the mass spectra profiles obtained for two different methods ($\mathbf{s_1}$ and $\mathbf{s_2}$) [50]:

$$cos\ \alpha = \frac{\mathbf{s}_1\mathbf{s}_2}{\|\mathbf{s_1}\|\|\mathbf{s_2}\|} \qquad \text{Eq. (9)}$$

2.4. Software

PARAFAC and PARAFAC2 were performed using PLS Toolbox 8.0.2 (Eigenvector Research inc, Wenatche, WA, USA) working under MATLAB R2015b (The Mathworks, Natick, MA, US). MCR-ALS analyses were carried out using the MCR-ALS 2.0 toolbox freely available at www.mcrals.info.

2.5. Abbreviations

TAGs have been abbreviated by means of three letters according to the three fatty acids bound to glycerol. The following abbreviations have been used: (P) palmitic acid (C16:0); (O) oleic acid (C18:1, Δ9); (S) stearic acid (C18:0); (L) linoleic acid (C18:2, Δ9,12); (A) arachidic acid (C20:0); (Po) palmitoleic acid (C16:1, Δ9); (Ln) α-linolenic acid (C18:3, Δ9,12,15). Cx:y indicates x number of carbons and y number of double bonds, and Δ indicates the position of these double bounds.

3. Results and discussion

An example of a 2D TIC LC×LC chromatogram obtained in the analysis of a corn oil sample is shown in Figure 2A. Seven important chromatographic regions were detected during the elution of TAGs in this LC×LC chromatogram. As an example, one of these seven regions was used to evaluate the trilinear behaviour of LC×LC-MS data. The selected region was number 2, zoomed in Figure 2B, which corresponded to time retention channels from 41 to 54 in the first-chromatographic dimension, and from 60 to 100 in the second chromatographic dimension (Figure 2A). Region 2 was the most interesting region of the chromatogram because two pairs of positional isomers were eluted in this region: SLO/SOL and PLO/POL. The SLO and SOL TAG pairs were separated in the first dimension column (peaks 1 and 2 in Figure 2B) and, therefore, their resolution was straightforward. However, PLO and POL co-eluted in both dimensions (peak 3 in Figure 2B), and their resolution was more troublesome. The size of the LC×LC-MS data cube, **D**, for this chromatographic region was 41 × 901 × 14 corresponding to 41 retention times of the second chromatographic dimension, 14 modulations from the first-

chromatographic dimension and 901 *m/z* values. When the data cube was unfolded into the $\mathbf{D}_{aug}$ matrix (see Figure 1B), the size of the matrix was 574 rows (41 × 14 retention times) and 901 columns (*m/z* values). Considering **D** and $\mathbf{D}_{aug}$, the study of the trilinear behaviour of LC×LC-MS data is described below.

Figure 2 near here

3.1. Study of LC×LC-MS data structure

In this section, the three-way structure of LC×LC-MS data is evaluated, in particular if the trilinear model is suitable for its investigation. The first test of this study consisted on the comparison of the SVD of three augmented data matrices containing the same data but arranged in three different ways (Figure S2A in the supplementary material). Experimental data was arranged in a column-wise augmented matrix $\mathbf{D}_{aug}$, as well as in a row-wise way (second dimension retention time in the common mode), and in a tube-wise way (each modulation in a row vector). If the trilinear model is accomplished, these three matrices should have the same chemical rank (mathematical rank in the absence of noise) and, therefore, their SVD analysis should give the same number of significant (not-noise) components [51]. Results showed that the number of significant components needed to explain the same amount of variance was lower for the column-wise data matrix, (larger explained variances) than for the other two augmented data matrices, row-wise and tube-wise, (they explain less variance) (Figure S2B in the supplementary material). These results indicated that the studied LC×LC-MS experimental data deviated from the trilinear model, and so the application of trilinear methods will be not appropriate for component resolution purposes.

Table 1 near here

To further study the effects of this deviation from the trilinear model, MCR-ALS, MCR-ALS trilinear, MCR-ALS trilinear allowing time shifting, PARAFAC and PARAFAC2 were applied and compared in the analysis of the selected chromatographic region (region 2 shown in Figure 2B). Table 1 shows the information related to the explained variance (R^2) and the lack of fit

(*LOF*) obtained by each of the models using three, four and five components. In the case of application of MCR-ALS, MCR-ALS trilinear, MCR-ALS trilinear allowing time shifting and PARAFAC2 methods, the finally selected number of components for further comparison was five. On the contrary, in the case of the PARAFAC method, the number of components was set to four because this was the model that gave the highest core consistency. When a data set does not fulfil the trilinear model requirements (see Chemometrics methodology section) for a particular number of components, methods based on the trilinear model will fit the data less efficiently (with a lower fit) and will give more unreasonable shape component profiles than other methods based on a bilinear model [35, 52]. Results presented in Table 1 show a clear gap between the *LOF* values of MCR-ALS and PARAFAC2 and the other three models (MCR-ALS fully or partially trilinear and PARAFAC). Moreover, MCR-ALS trilinear allowing time shifting gave better values of *LOF* and R^2 than MCR-ALS trilinear but worse than bilinear MCR-ALS. This result indicated that not only retention times, but also peak shapes changed among modulations, and consequently the analysed LC×LC-MS data could not be considered as trilinear. An example of the differences in elution profiles resolved by MCR-ALS, MCR-ALS trilinear allowing time shifting and MCR-ALS trilinear is shown in Figure S3 in supplementary material.

Furthermore, core consistency diagnostic can also be used to evaluate the trilinear behaviour of a data set [53]. When the evaluated data cannot be described by a trilinear model or too many components are used in the model, core consistency will differ from 100% [53]. In our case, the obtained results for the PARAFAC model gave a core consistency of 80% when 4 components were used, and when the model was performed using three or five components the resulted core consistency decreased significantly (approximately to 65%). All these results indicated that the analysed LC×LC-MS data did not accomplish the trilinear model requirements. Most probably, this deviation from the trilinearity is caused in part by changes in retention times of the same peak on the second dimension column modulation (shifting) and in another part by changes in

peak shapes. A short explanation of PARAFAC2 limitations is given below that agrees with those arguments described in reference [29].

Figure 3 near here

In order to complete the comparison between these methods, the resolved elution profiles and mass spectra by each resolution method were contrasted. Figure 3 shows the obtained elution profiles in both chromatographic dimensions after applying the five methods. Considering the shape of the profiles recovered by the different methods, the obtained profiles may look reasonably good from a chemical point of view. Only the second column profiles resolved by PARAFAC2 (Figure 3E) were an exception. Since in PARAFAC2 non-negativity constraints could not be applied to second dimension chromatographic mode, the profiles resolved by this method were worse from a physical point of view than those provided by PARAFAC, MCR-ALS, MCR-ALS trilinear methods allowing time shifting or MCR-ALS trilinear. This outcome agreed with the results obtained by Bortolato and Olivieri in the comparison of MCR-ALS and PARAFAC2 for chromatographic analysis [29]. In their work, Bortolato and Olivieri exposed PARAFAC2 limitations and the main reasons for its failure. Briefly, the fact that PARAFAC2 requires that the cross-product of different $\mathbf{X}_k$ should be equal in all modulations implies two important consequences: (1) peak shape should be the same in all modulations for every component *n*, (2) peak shifts are only tolerated in PARAFAC2 for non-coeluting components and in the absence of interferences, otherwise the cross-product constant condition assumed by PARAFAC2 is not fulfilled anymore. To further study retention time shifting and changes in peak shapes, SVD analysis of the resolved elution profiles was performed (Table S1). If the studied LC×LC-MS data were trilinear, then retention time and peak shape of a chemical compound would be equal in all modulations. Consequently, the SVD of the resolved elution profile of an individual compound would give only one significant component. Since MCR-ALS trilinear forced the resolved elution profiles to be constant over all modulations, the SVD of resolved elution profiles gave only one relevant singular component in all cases. Otherwise, the SVD of SLO, POL, SOL and PLO elution profiles resolved by MCR-ALS bilinear gave 6,

5, 10, and 9 significant components respectively, which clearly indicated that these profiles were not trilinear. Moreover, when the SVD was performed on SLO, POL, SOL and PLO elution profiles resolved by MCR-ALS trilinear allowing time shifting, the obtained number of significant components was 3, 4, 4 and 7 respectively. All these results demonstrated that even allowing time shifting, more than one component was necessary to explain the elution profile of the same component over all modulations, indicating the non-fulfilment of the trilinear model for this component. Moreover, the lack of trilinearity of the considered data was not only due to retention time shifting but also to changes in peak shapes between consecutive modulations.

Table 2 near here

Finally, comparison of the resolved mass spectra for each one of the pure components (SLO, SOL, POL and PLO) obtained by each method is shown in Table 2. Results of this comparison are shown in terms of the angle between the vectors defined by the resolved mass spectra for each pair of methods. Most of these comparisons showed a rather good matching between profiles, with angles lower than 2 degrees. However, the resolution of the PLO isomer showed angles higher than 20 degrees between MCR-ALS and the other methods. This compound was the most difficult to resolve because it was completely co-eluted with POL and only the MCR-ALS bilinear method properly resolved PLO.

Taking into account all the results obtained until now, the analysed LC×LC-MS data were confirmed not to have behaviour consistent with a trilinear model. Consequently, the use of the PARAFAC method would require a preliminary peak alignment procedure as the one suggested by Allen and Rutan to deal with LC×LC-DAD data [36] and, probably, would still have problems with the changes in peak shape changes in case of coelution. On the other hand, MCR-ALS or PARAFAC2 did not require the initial peak alignment step, which can be considered to be an advantage. Finally, PARAFAC2 was not able to obtain reliable second dimension elution profiles because of the application of constraints to this mode were not allowed and in many circumstances. So, PARAFAC2 did not solve data appropriately due to

peak shape changes among modulations. On the contrary, MCR-ALS was able to resolve LC×LC-MS data properly and allowed an easy interpretation of the achieved results.

3.2. Application of MCR-ALS to LC×LC-MS data

The potential application of MCR-ALS to LC×LC-MS data is discussed in this section for the resolution of TAGs isomers. With this aim, the analysis of chromatographic regions number 1, 2 and 3 (see Figure 2A) is described below. Regions 1 and 3 (Zoomed in Figure S4 of the supplementary material) were selected because these regions contained a pair of chain isomers that co-eluted in both chromatographic dimensions and could not be resolved by traditional means. In the case of chain isomers, mass spectra show the same protonated molecular ion for both isomers and, also, some diacylglycerol fragments are the same. Region 1 corresponds to retention time channel from 30 to 42 in the first-chromatographic dimension and from 55 to 110 in the second chromatographic dimension. Region 3 corresponds to retention time channel from 61 to 70 in the first-chromatographic dimension and from 50 to 90 in the second chromatographic dimension. The resolution of region number 2 was especially challenging due to the elution of two pairs of positional isomers. In this case both, the protonated molecular ion and all diacylglycerol fragments were the same for both positional isomers, which could be only distinguished by the relative abundance of their fragments. The chromatographic characteristics of this region have already been described in the previous section.

3.2.1. MCR-ALS resolution TAGs positional isomers

Figure 4 near here

The resolution of the two pairs of positional isomers contained in region number 2 shows the advantages of the application of MCR-ALS to LC×LC-MS data. As explained in the previous section, the MCR-ALS model with five components was selected. The percentage of explained data variance (R^2) was 99.9% and the *LOF* was 3.1%. Four of these components (see below) were assigned to TAGs positional isomers (SOL, SLO, POL and PLO) whereas the fifth component was assigned to a baseline contribution. In this last case, both chromatographic

profiles and resolved mass spectrum showed profiles that could not be associated with a compound with chemical meaning.

Figure 4, shows the results obtained after MCR-ALS resolution of region 2. Figure 4A depicts the first-column elution profiles of the resolved components (baseline component is also shown). As it can be observed, SLO (blue profile) was separated from the other three compounds in the first-column (argentation chromatography) due to the interaction of Ag(I) ions with the double bonds, so its identification was straightforward. Figure 4B corresponds to the second dimension column (reverse phase C18) elution profiles, where it can be seen that SOL (black) was separated from POL (red) and PLO (green) due to the difference in their partition numbers (48 for SOL and 46 for POL and PLO). Therefore, SOL identification was also possible. On the contrary, PLO (green) was embedded in PLO (red), what made their separation and identification more difficult. Figure 5A shows the resolved mass spectra of the four components, which were used to identify the TAGs by comparison with their experimental reference mass spectra. TAGs mass spectra were characterised by the mass of the protonated molecular ion ($[M+H]^+$) and the mass of all possible protonated diacylglycerols fragments ($[DG+H]^+$). As it is shown in this Figure, the two TAGs positional isomers have exactly the same *m/z* values for the masses of the $[M+H]^+$ and three $[DG+H]^+$ ions and their mass spectra only differ in the relative abundance of the three different fragments of $[DG+H]^+$. Moreover, since the two positional isomers POL and PLO are eluted at the same retention time, using traditional and commercial chromatographic software based on individual *m/z* signals of various ions their distinction is really difficult and usually fail in their qualitative and quantitative determination. Outstandingly, this difficult problem could be solved by the proposed MCR-ALS method, which was capable of resolving and identifying separately these two isomers. This fact is highly relevant in the study of positional isomers.

Figure 5 near here

As an example of how TAGs identification was carried out, inserts in Figure 5A show the zoomed diacylglycerols mass region (from 565 to 610 *m/z*) of the resolved MCR-ALS mass spectra for the four TAGs (SOL, SLO, POL and PLO) and Figure 5B depicts their experimental reference mass spectra obtained from references [10] and [54]. Fragmentation of the POL and PLO isomers gives three possible diacylglycerols fragments ($[DG+H]^+$): LO (601.6 *m/z*), LP (575.5 *m/z*) and PO (577.5 *m/z*). The relative abundance of these three diacylglycerol fragments depends on the different probability that ester bonds of TAGs break. The fatty acid least likely to split off is the most hindered one, i.e. at position 2 of the glycerol. Consequently, for the POL isomer the signal of LP (breaking O-ester bond, at 575.5 *m/z*) is the least abundant (ratio PO/PL = 1.39) whereas for the PLO isomer the PO (breaking L ester bond, at 577.5 *m/z*) has the lowest abundance (ratio PO/PL = 0.71). The main difference between the mass spectra of PLO and POL isomers lies only in the relative abundance of their signals at 575.5 and 577.5 *m/z*. In Figure 5A, it can be seen that in the resolved mass spectrum for POL the signal intensity of 577.5 *m/z* (0.8 a.u.) was higher than the signal intensity at 575.5 *m/z* (0.6 a.u.), as happen in its experimental reference spectra (Figure 5B). In the case of the resolved mass spectrum for PLO (Figure 5A), the signal intensity at 575.5 *m/z* (0.8 a.u.) was higher than the signal intensity of 577.5 *m/z* (0.55 a.u.), in the same way than in the experimental reference spectra (Figure 5B). This shows again the great potential of MCR resolution compared to traditional means where these two species would be extremely difficult to be resolved when they are coeluting. SLO and SOL isomers were identified using the same described strategy.

3.2.2. MCR-ALS resolution of TAGs chain isomers

Figure 6 near here

Region 1 of Figure 2A was resolved by an MCR-ALS model with nine components. The percentage of explained variance (R^2) was 99.5% and *LOF* was 6.8%. Eight of these components were assigned to TAGs (SLS, ALP and other TAGs whose resolution was not difficult) and the ninth component was assigned to a baseline contribution. Figure 6 shows the resolved elution profiles and mass spectra of SLS (MCR-ALS resolved component 3) and ALP

(MCR-ALS resolved component 7). Figure 6A depicts the resolved elution profiles, where SLS and ALP were strongly co-eluting in both chromatographic dimensions, making their resolution challenging. Figure 6B shows the MCR-ALS resolved mass spectra of SLS and ALP and Figure 6C depicts a zoom of the diacylglycerols mass region (from 570 to 640 *m/z*) of these two spectra. SLS and ALP chain isomers have exactly the same *m/z* value of the protonated molecular ion $[M+H]^+$ (887.5 *m/z*) and of one of the diacylglycerol fragments, $[DG+H]^+$ (607.5 *m/z* for both AP and SS). However, they also differ in the *m/z* values of other of the two diacylglycerols fragments (575.6 *m/z* for LP, 631.5 *m/z* for AL and 603.6 *m/z* for LS in Figure 6) [10]. The main difference between the mass spectra of ALP and SLS lies in the presence of the signals at 575.6 and 631.5 *m/z*, which should appear only in ALP mass spectrum and of the signal at 603.6, which should appear only in SLS mass spectrum. In Figure 6C, it can be seen that signal intensities at 575.6 and 631.6 were low in the MCR-ALS resolved mass spectrum of SLS, whereas in the MCR-ALS resolved mass spectrum of ALP these signals showed a high intensity. Moreover, the signal at 603.6 *m/z* is only present in SLS mass spectrum. Therefore, the two chain isomers could be identified.

Finally, a new MCR-ALS model with five components was employed to resolve region 3 of Figure 2. The percentage of explained variance (R^2) was 99.8% and the *LOF* was 3.2%. Four of these components were assigned to TAGs (POLn, PoOL and to other TAGs whose resolution was rather simple with no coelution), and the fifth component was again assigned to a noisy background contribution. Figure S5 in supplementary material shows the resolved elution profiles and mass spectra of PoOL (MCR-ALS resolved component 1) and POLn (MCR-ALS resolved component 4), two chain isomers. The resolved elution profiles are represented on Figure S5A, where can be observed that the two chain isomers were heavily coeluting, with POLn totally embedded in PoOL which hindered their identification. Figure S5B shows the MCR-ALS resolved mass spectra of PoOL and POLn and Figure S5C gives the zoomed *m/z* region where these two diacylglycerols have their signals (from 555 to 625 *m/z*). PoOL and POLn chain isomers have exactly the same *m/z* value of the masses of $[M+H]^+$ (855.6 *m/z*) and

one of the fragments $[DG+H]^+$ (573.5 *m/z* for PoL and PLn) but they differ in the values of the other two diacylglycerols fragments (575.5 *m/z* for PoO, 603.5 *m/z* for LO , 577.5 for PO and 599.5 *m/z* for OLn) [10]. Figure S5C shows that the main differences observed in the resolved mass spectra of PoOL and POLn were in the presence of the signal at 575.5 *m/z* in PoOL mass spectrum and of the signals at 577.5 *m/z* and 599.5 *m/z* in the POLn one. Therefore, this is again a clear example of the potential of using the MCR-ALS proposed approach to resolve and distinguish these two extremely similar chain isomers.

Conclusions

A comprehensive study about the potential of using the MCR-ALS method for the analysis of LC×LC-MS data has been performed. First, it has been proven that LC×LC-MS data does not fulfil the frequently postulated trilinearity requirements since both large retention time shifts and changes in peak shapes between consecutive modulations are occurring in general. Thus, methods based on bilinear models are more suitable to analyse LC×LC-MS data than trilinear model based methods like PARAFAC or PARAFAC2. Using trilinear model based methods, peak retention time shifts and shape changes between modulations and therefore peak alignment procedures are needed. It is proposed therefore the general use of the MCR-ALS method to take full advantage of LC×LC-MS data in the analysis of complex natural samples where strong coelutions and spectra overlap (isomeric species) are frequently encountered. MCR-ALS strategy is shown to be rather simple, and results can be easily interpreted from the directly resolved elution profiles in the two chromatographic dimensions and from the resolved mass spectra of every constituent contained in the mixture sample. Using these profiles, qualitative (identification) information can be easily derived and used for interpretation. Different examples are shown and described in detail in the analysis of different co-eluting triacylglycerols structural isomers in vegetable oils. For instance, POL and PLO positional isomers were resolved and identified although they were coeluting completely in both chromatographic dimensions.

Acknowledgements

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 n. 320737. Also, recognition from the Catalan government (grant 2014 SGR 1106) is acknowledged.

Conflict of interest statement

The authors declare that they have no competing interests.

References

[1] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E.G. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, J. Chromatogr. A. 1168 (2007) 3-43.
[2] R.A. Shellie, P.R. Haddad, Comprehensive two-dimensional liquid chromatography, Anal. Bioanal. Chem. 386 (2006) 405-415.
[3] K.M. Pierce, B. Kehimkar, L.C. Marney, J.C. Hoggard, R.E. Synovec, Review of chemometric analysis techniques for comprehensive two dimensional separations data, J. Chromatogr. A. 1255 (2012) 3-11.
[4] K.M. Pierce, J.C. Hoggard, R.E. Mohler, R.E. Synovec, Recent advancements in comprehensive two-dimensional separations with chemometrics, J. Chromatogr. A. 1184 (2008) 341-352.
[5] J.T.V. Matos, S. Freire, R. Duarte, A.C. Duarte, Profiling water-soluble organic matter from urban aerosols usingcomprehensive two-dimensional liquid chromatography, Aerosol Sci. Technol. 49 (2015) 381-389.
[6] S. Julka, H. Cortes, R. Harfmann, B. Bell, A. Schweizer-Theobaldt, M. Pursch, L. Mondello, S. Maynard, D. West, Quantitative characterization of solid epoxy resins using comprehensive two dimensional liquid chromatography coupled with electrospray ionization-time of flight mass spectrometry, Anal. Chem. 81 (2009) 4271-4279.
[7] M. Lísa, E. Cífková, M. Holčapek, Lipidomic profiling of biological tissues using off-line two-dimensional high-performance liquid chromatography-mass spectrometry, J. Chromatogr. A. 1218 (2011) 5146-5156.
[8] Q. Yang, X. Shi, Y. Wang, W. Wang, H. He, X. Lu, G. Xu, Urinary metabonomic study of lung cancer by a fully automatic hyphenated hydrophilic interaction/RPLC-MS system, J. Sep. Sci. 33 (2010) 1495-1503.
[9] H.P. Bailey, S.C. Rutan, Comparison of chemometric methods for the screening of comprehensive two-dimensional liquid chromatographic analysis of wine, Anal. Chim. Acta. 770 (2013) 18-28.
[10] E.J.C. van der Klift, G. Vivó-Truyols, F.W. Claassen, F.L. van Holthoon, T.A. van Beek, Comprehensive two-dimensional liquid chromatography with ultraviolet, evaporative light scattering and mass spectrometric detection of triacylglycerols in corn oil, J. Chromatogr. A. 1178 (2008) 43-55.
[11] J.M. Bosque-Sendra, L. Cuadros-Rodríguez, C. Ruiz-Samblás, A.P. de la Mata, Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data-A review, Anal. Chim. Acta. 724 (2012) 1-11.

[12] L.W. Hantao, H.G. Aleme, M.P. Pedroso, G.P. Sabin, R.J. Poppi, F. Augusto, Multivariate curve resolution combined with gas chromatography to enhance analytical separation in complex samples: A review, Anal. Chim. Acta. 731 (2012) 11-23.
[13] S. Mas, A. de Juan, R. Tauler, A.C. Olivieri, G.M. Escandar, Application of chemometric methods to environmental analysis of organic pollutants: A review, Talanta. 80 (2010) 1052-1067.
[14] Y.P. Lin, D.Y. Si, C.X. Liu, Advances of liquid chromatography/mass spectrometry combined with chemometric approaches applied to metabonomics, Chin. J. Anal. Chem. 35 (2007) 1535-1540.
[15] A.E. Sinha, C.G. Fraga, B.J. Prazen, R.E. Synovec, Trilinear chemometric analysis of two-dimensional comprehensive gas chromatography-time-of-flight mass spectrometry data, J. Chromatogr. A. 1027 (2004) 269-277.
[16] A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, Algorithm for locating analytes of interest based on mass spectral similarity in GC × GC-TOF-MS data: Analysis of metabolites in human infant urine, J. Chromatogr. A. 1058 (2004) 209-215.
[17] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700-2709.
[18] L.W. Hantao, B.R. Toledo, F.A. De Lima Ribeiro, M. Pizetta, C.G. Pierozzi, E.L. Furtado, F. Augusto, Comprehensive two-dimensional gas chromatography combined to multivariate data analysis for detection of disease-resistant clones of Eucalyptus, Talanta. 116 (2013) 1079-1084.
[19] H. Parastar, J.R. Radović, M. Jalali-Heravi, S. Diez, J.M. Bayona, R. Tauler, Resolution and quantification of complex mixtures of polycyclic aromatic hydrocarbons in heavy fuel oil sample by means of GC × GC-TOFMS combined to multivariate curve resolution, Anal. Chem. 83 (2011) 9289-9297.
[20] H. Parastar, M. Jalali-Heravi, R. Tauler, Comprehensive two-dimensional gas chromatography (GC×GC) retention time shift correction and modeling using bilinear peak alignment, correlation optimized shifting and multivariate curve resolution, Chemometrics Intell. Lab. Syst. 117 (2012) 80-91.
[21] H. Parastar, J.R. Radović, J.M. Bayona, R. Tauler, Solving chromatographic challenges in comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry using multivariate curve resolution-alternating least squares, Anal. Bioanal. Chem. 405 (2013) 6235-6249.
[22] H. Parastar, R. Tauler, Multivariate curve resolution of hyphenated and multidimensional chromatographic measurements: A new insight to address current chromatographic challenges, Anal. Chem. 86 (2014) 286-297.
[23] J.R. Radović, K.V. Thomas, H. Parastar, S. Díez, R. Tauler, J.M. Bayona, Chemometrics-assisted effect-directed analysis of crude and refined oil using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Environ. Sci. Technol. 48 (2014) 3074-3083.
[24] 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.
[25] 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. Chemometr. 26 (2012) 474-486.
[26] H.P. Bailey, S.C. Rutan, Chemometric resolution and quantification of four-way data arising from comprehensive 2D-LC-DAD analysis of human urine, Chemometrics Intell. Lab. Syst. 106 (2011) 131-141.
[27] P.G. Stevenson, M. Mnatsakanyan, G. Guiochon, R.A. Shalliker, Peak picking and the assessment of separation performance in two-dimensional high performance liquid chromatography, Analyst. 135 (2010) 1541-1550.

[28] G. Vivó-Truyols, Bayesian approach for peak detection in two-dimensional chromatography, Anal. Chem. 84 (2012) 2622-2630.
[29] S.A. Bortolato, A.C. Olivieri, Chemometric processing of second-order liquid chromatographic data with UV-vis and fluorescence detection. A comparison of multivariate curve resolution and parallel factor analysis 2, Anal. Chim. Acta. 842 (2014) 11-19.
[30] A.W. Bowman, Applied smoothing techniques for data analysis : the kernel approach with S-Plus illustrations, in: A. Azzalini (Ed.), Clarendon Press ;, Oxford :, 1997.
[31] R. Bro, PARAFAC. Tutorial and applications, Chemometrics Intell. Lab. Syst. 38 (1997) 149-171.
[32] J.C. Hoggard, R.E. Synovec, Automated resolution of nontarget analyte signals in GC x GC-TOFMS data using parallel factor analysis, Anal. Chem. 80 (2008) 6677-6688.
[33] R. Bro, C.A. Andersson, H.A.L. Kiers, PARAFAC2 - Part II. Modeling chromatographic data with retention time shifts, J. Chemometr. 13 (1999) 295-309.
[34] H.A.L. Kiers, J.M.F. Ten Berge, R. Bro, PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model, J. Chemometr. 13 (1999) 275-294.
[35] A. De Juan, R. Tauler, Comparison of three-way resolution methods for non-trilinear chemical data sets, J. Chemometr. 15 (2001) 749-772.
[36] R.C. Allen, S.C. Rutan, Semi-automated alignment and quantification of peaks using parallel factor analysis for comprehensive two-dimensional liquid chromatography-diode array detector data sets, Anal. Chim. Acta. 723 (2012) 7-17.
[37] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemometrics Intell. Lab. Syst. 140 (2015) 1-12.
[38] A. de Juan, R. Tauler, Factor analysis of hyphenated chromatographic data. Exploration, resolution and quantification of multicomponent systems, J. Chromatogr. A. 1158 (2007) 184-195.
[39] A. de Juan, J. Jaumot, R.A. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods 6 (2014) 4964-4976.
[40] A. de Juan, R. Tauler, Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications, Crit. Rev. Anal. Chem. 36 (2006) 163-176.
[41] G.H. Golub, C.F.V. Loan, Matrix computations, third ed., Johns Hopkins University Press, Baltimore, 1996.
[42] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425-1432.
[43] W. Windig, D.A. Stephenson, Self-modeling mixture analysis of second-derivative near-infrared spectral data using the simplisma approach, Anal. Chem. 64 (1992) 2735-2742.
[44] A. De Juan, S.C. Rutan, R. Tauler, Two-Way Data Analysis: Multivariate Curve Resolution - Iterative Resolution Methods, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier, Oxford, 2009, pp. 325-344.
[45] R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate Curve Resolution, in: S.D. Brown, R. Tauler, B. Walczak (Eds.) Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Elsevier, Oxford, 2009, pp. 473-505.
[46] J.M. Amigo, T. Skov, R. Bro, ChroMATHography: Solving chromatographic issues with mathematical models and intuitive graphics, Chem. Rev. 110 (2010) 4582-4605.
[47] N.M. Faber, R. Bro, P.K. Hopke, Recent developments in CANDECOMP/PARAFAC algorithms: A critical review, Chemometrics Intell. Lab. Syst. 65 (2003) 119-137.
[48] K.S. Booksh, Z. Lin, Z. Wang, B.R. Kowalski, Extension of trilinear decomposition method with an application to the flow probe sensor, Anal. Chem. 66 (1994) 2561-2564.
[49] N.M. Faber, Towards a rehabilitation of the generalized rank annihilation method (GRAM), Anal. Bioanal. Chem. 372 (2002) 683-687.
[50] J. Jaumot, J.C. Menezes, R. Tauler, Quality assessment of the results obtained by multivariate curve resolution analysis of multiple runs of gasoline blending processes, J. Chemometr. 20 (2006) 54-67.
[51] R. Tauler, Multivariate curve resolution applied to second order data, Chemometrics Intell. Lab. Syst. 30 (1995) 133-146.

[52] A. De Juan, S.C. Rutan, R. Tauler, D. Luc Massart, Comparison between the direct trilinear decomposition and the multivariate curve resolution-alternating least squares methods for the resolution of three-way data sets, Chemometrics Intell. Lab. Syst. 40 (1998) 19-32.
[53] R. Bro, H.A.L. Kiers, A new efficient method for determining the number of components in PARAFAC models, J. Chemometr. 17 (2003) 274-286.
[54] P. Dugo, O. Favoino, P.Q. Tranchida, G. Dugo, L. Mondello, Off-line coupling of non-aqueous reversed-phase and silver ion high-performance liquid chromatography-mass spectrometry for the characterization of rice oil triacylglycerol positional isomers, Journal of Chromatography A 1041 (2004) 135-142.

Figure Captions

Figure 1. Graphical description of two possible LC×LC-MS data arrangements. A) LC×LC-MS data arranged in a data cube showing a matrix for each modulation, a mass spectrum for each row of the $\mathbf{D_K}$ matrices and a 2D chromatogram if the cube is considered by the side. B) LC×LC-MS data arranged in a column-wised (in the *m/z* mode) augmented data matrix.

Figure 2. A) Bidimensional chromatogram of the corn oil sample. Seven important chromatographic regions are marked and numbered. B) Zoomed view of region of interest number 2 with the three important peaks numbered (peak 1 is SOL, peak 2 is SLO and peak 3 corresponds to POL and PLO). Colorbar indicates peak intensity.

Figure 3. First- (left) and second- (right) chromatographic dimension elution profiles recovered with: A) MCR-ALS; B) MCR-ALS trilinear allowing time shifting; C) MCR trilinear; D) PARAFAC; and E) PARAFAC2. Colours indicate the four TAGs: Blue is SLO, black is SOL, red is POL and green is PLO.

Figure 4. MCR-ALS resolved elution profiles for chromatographic region 2. A) MCR-ALS resolved first-column elution profiles. B) MCR-ALS resolved second column elution profiles.

Figure 5. A) MCR-ALS resolved pure mass spectra profiles for SLO, SOL, POL and PLO. Inserts show the diacylglycerols mass region (from 565 to 610 *m/z*) of each TAG. B) Reference model spectra of the TAGs used for their identification. SOL and SLO spectra adapted from [54] and POL and PLO adapted from [10].

Figure 6. MCR-ALS resolution of SLS and ALP chain isomers. A) Resolved elution profiles in the fourteen modulations taken from the first dimension column. B) Resolved mass spectra. C) MCR-ALS resolved mass spectra zoomed in the diacylglycerols mass region (from 570 to 640 *m/z*).

Table 1. Comparison of results obtained by the application of MCR-ALS, MCR-ALS trilinear allowing time shifting, MCR-ALS trilinear, PARAFAC and PARAFAC2 methods

Method	Number of Components	R^2 (%)[1]	*LOF* (%)[2]	Number of ALS iterations	Core Consistency (%)	Selected for comparison[3]
MCR-ALS	3	99.6	6	18	-	
	4	99.8	4	17	-	
	5	99.9	3	23	-	*
MCR-ALS TRILINEAR ALLOWING TIME SHIFTING	3	97.0	18	3	-	
	4	97.0	17	3	-	
	5	97.1	17	4	-	*
MCR-ALS TRILINEAR	3	89.3	33	2	-	
	4	89.3	32	6	-	
	5	95.3	21	11	-	*
PARAFAC	3	96.2	19	61	67	
	4	97.9	14	51	80	*
	5	98.6	11	23	65	
PARAFAC 2	3	99.6	7	1295	99	
	4	99.8	4	918	99	
	5	99.9	3	11287	99	*

[1] Calculated according to equation 7.

[2] Calculated according to equation 8.

[3] Indicates the model chosen to compare the five methods.

Table 2. Data matching between the resolved mass spectra of SLO, SOL, POL and PLO in region 2 by MCR-ALS, MCR-ALS trilinear allowing time shifting, MCR-ALS trilinear PARAFAC and PARAFAC2.

	Angle (α)			
	SLO	SOL	POL	PLO
MCR-ALS VS MCR-ALS TRILINEAR	0.6	1.9	0.5	**34**
MCR-ALS VS MCR-ALS TRILINEAR ALLOWING TIME SHIFTING	0.2	1.1	1.6	**23**
MCR-ALS VS PARAFAC	0.7	4.5	1.0	**35**
MCR-ALS VS PARAFAC2	0.3	2.3	1.0	**34**
MCR-ALS TRILINEAR ALLOWING TIME SHIFTING VS MCR-ALS TRILINEAR	0.4	2.4	1.6	**38**
MCR-ALS TRILINEAR ALLOWING TIME SHIFTING VS PARAFAC	0.9	3.9	1.8	**39**
MCR-ALS TRILINEAR ALLOWING TIME SHIFTING VS PARAFAC2	0.2	3.1	1.8	**25**
MCR-ALS TRILINEAR VS PARAFAC	1.2	5.3	0.8	1.3
MCR- TRILINEAR VS PARAFAC 2	0.4	1.9	0.6	0.6
PARAFAC VS PARAFAC 2	1.0	6.1	0.9	0.6

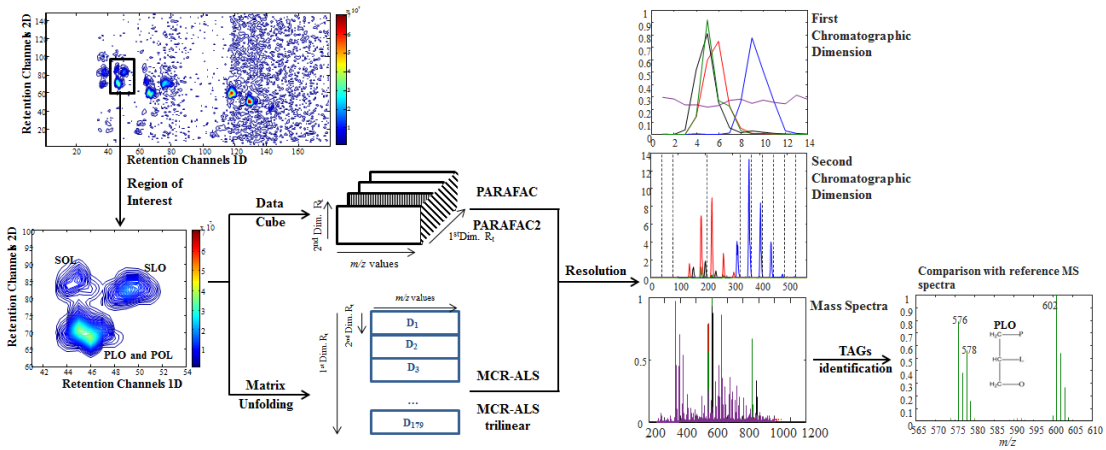
Highlights

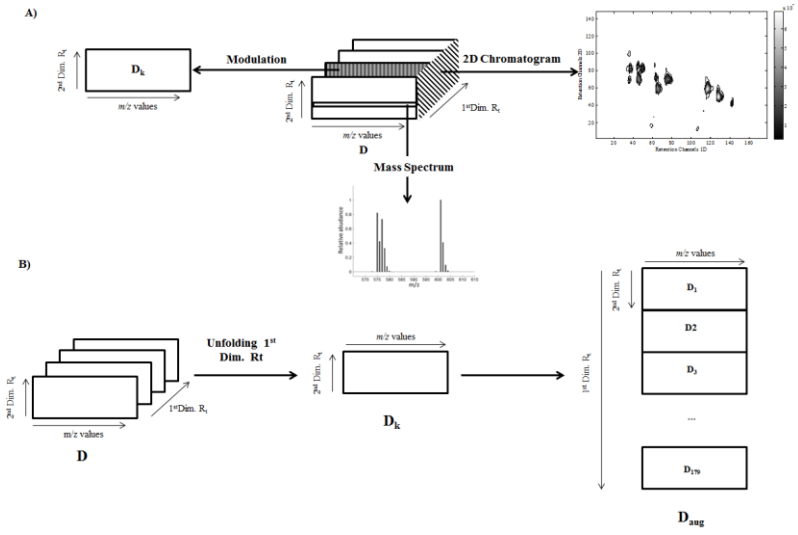
A multivariate curve resolution based approach is proposed to resolve LC×LC-MS data.

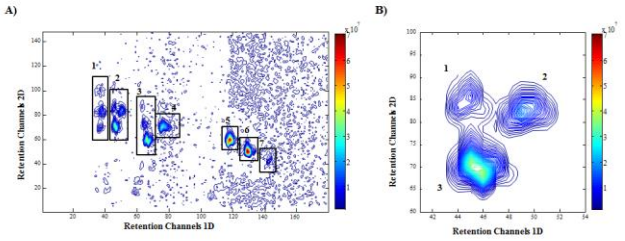
Bilinear and trilinear behavior of LC×LC-MS data is discussed.

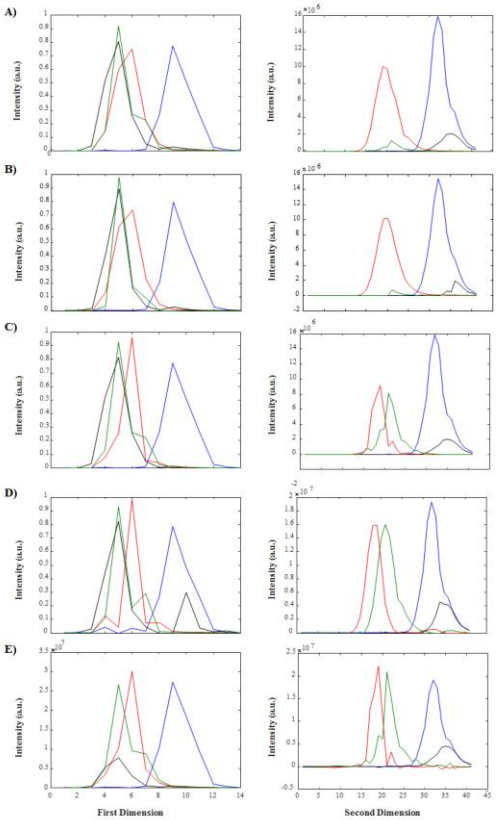
Results obtained by MCR-ALS, MCR-ALS trilinear, PARAFAC and PARAFAC2 are compared.

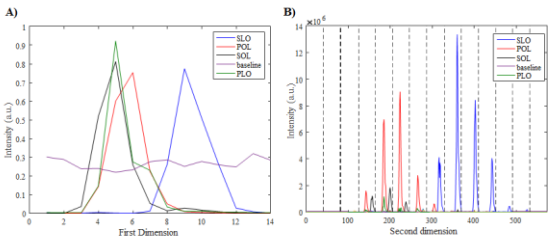
TAGs positional isomers in corn oil are resolved and identified.


Retention Channels 2D
Retention Channels 1D
Region of Interest
SOL
SLO
PLO and POL
Data Cube
Matrix Unfolding
PARAFAC
PARAFAC2
MCR-ALS
MCR-ALS trilinear
Resolution
First Chromatographic Dimension
Second Chromatographic Dimension
Mass Spectra
TAGs identification
Comparison with reference MS spectra
PLO


A)
Modulation
2D Chromatogram
Mass Spectrum
B)
Unfolding 1st Dim. Rt
D
Dk
Daug


A)
B)
Retention Channels 2D
Retention Channels 1D
1
2
3
4
5
6


A)
B)
C)
D)
E)
Intensity (a.u.)
First Dimension
Second Dimension


A)
B)
SLO
POL
SOL
baseline
PLO
Intensity (a.u.)
First Dimension
Second dimension