Exploration of liquid chromatographic-diode array data for Argentinean wines by extended multivariate curve resolution

Pablo L. Pisano,[a] María F. Silva,[b] Alejandro C. Olivieri[a,*]

[a] *Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina*

[b] *Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo, Instituto de Biología Agrícola de Mendoza (IBAM-CONICET), Alte. Brown 500, Chacras de Coria, 5505, Mendoza, Argentina*

* Corresponding author. Tel/fax: +54 341 4372704. E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri)

ABSTRACT

Second-order data were measured using high-performance liquid-chromatography with diode array detection (HPLC-DAD) for a number of wine samples, which were directly injected in the HPLC-DAD system without sample pre-treatment. The data were arranged in data matrices whose modes were elution time and UV-visible absorption wavelength, and processed by extended multivariate curve resolution coupled to alternating least-squares (MCR-ALS). The individual data matrices were organized in a row-wise augmented data matrix sharing the time subspace, due to the high spectral similarity among several sample components. This required previous time alignment of the chromatograms using a suitable synchronization algorithm, in order to produce a bilinear augmented data matrix to be processed by MCR-ALS. The latter algorithm led to resolved component chromatograms and spectra, from which component scores could be estimated, which are proportional to the relative component concentrations in each studied sample. The matrix of sample scores were then submitted to principal component analysis, which was applied for data exploration according to grape *varietal* and geographical origin. The results showed that the present data generation and analysis is useful for the discrimination of all samples of the Malbec *varietal* from the remaining ones, but achieved partial success regarding geographical origin.

Keywords: Liquid chromatography; Multivariate curve resolution; Principal component analysis; Direct injection; Wine data exploration.

1. Introduction

Wine is a complex matrix composed of water, ethanol and a variety of chemical compounds such as peptides, proteins, carbohydrates, thiols, and phenolic compounds [1]. The latter ones can be classified into flavonoids (flavanols, flavonols, dihydroflavonols, and anthocyanins) and non-flavonoids (phenolic acids, phenols, and stilbenes) [2]. Flavonoids share a common skeleton consisting of two phenolic rings (A and B) linked by a heterocyclic pyran ring (C), as shown in Fig. 1. Anthocyanins and flavanols are particularly abundant in grape and wine and are essential to wine quality. Indeed, anthocyanins are the red pigments of grapes and are responsible for the colour of red wines, whereas flavanols contribute to taste (especially astringency and bitterness) [3]. Due to the presence of aromatic rings in their structure, most phenolic compounds present in wine absorb UV-visible radiation with an absorption maximum at 280 nm, with the exception of anthocyanins (520 nm), flavonols (360 nm) and phenolic acids (320 nm) [2].

Due to the complexity of wine data obtained by usual instrumental techniques, it is not possible to resolve or quantify all the chemical constituents present in wine. Therefore, the combination of these techniques with chemometric analysis can reveal latent patterns in the data, which may enable classification of the samples in terms of *varietal*, geographical origin, aging, adulteration, etc. [4]. Several instrumental techniques have been employed for wine classification, such as gas chromatography-mass spectrometry (GC-MS) [5-7], high-performance liquid chromatography with diode array detection (HPLC-DAD) [8,9] or liquid chromatography coupled to mass spectrometric detection (LC-MS) [10-12], proton nuclear magnetic resonance (^{1}H NMR) [13,14], near-infrared spectroscopy (NIR) [15,16], capillary electrophoresis (CE) [17,18] and elemental analysis [19,20]. To achieve sample discrimination, the obtained data have been processed by different chemometric algorithms such as principal component analysis (PCA), linear discriminant analysis (LDA), partial least-

squares-discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), artificial neural networks (ANN), etc. [4].

In the past few years, several reports employed HPLC-DAD coupled chemometric tools in order to classify wines [21-26]. Nevertheless, to our knowledge, there are no reports of wine classification by direct injection HPLC-DAD without sample pre-treatment coupled to multivariate curve resolution-alternating least-squares (MCR-ALS) as data processing algorithm. In this work we employed the latter combination of techniques to attempt classification of wines by grape *varietal* and geographical origin of some Argentinean wines. The application of the MCR-ALS algorithm is usually made by joining the elution time-spectral data matrices adjacent to each other sharing the spectral subspace (i.e., by column-wise augmentation), creating the so-called augmented data matrix before MCR-ALS decomposition. However, for reasons which will be clear below, we adopted the somewhat exceptional procedure of augmentation by sharing the time subspace (i.e., row-wise augmentation) [27,28]. This required previous alignment of the chromatographic-spectral data matrix in order to alleviate the time shifts between runs [29].

The purpose of the present work is to model direct injection LC-DAD data for wine samples with MCR-ALS, in order to extract information which may allow for wine discrimination according to *varietal* and geographical origin. The results of this data exploration indicate that the Malbec *varietal* can be adequately discriminated from the remaining ones, while only partial success is obtained regarding the geographical origin of samples.

2. Experimental section

2.1. Reagents and standards

HPLC grade acetonitrile were purchased from Panreac (Barcelona, Spain), formic acid from Cicarelli (Rosario, Argentina) was pro analysis grade and used directly. Ultrapure water (18.2 MΩ cm) was obtained from a Milli-Q water purification system (Millipore Corp., Bedford, USA).

2.2. Wine Samples

The 27 wine samples were obtained from red grapes of *V. Vinífera* L. of eight varieties [Aspiran (A), Bonarda (B), Cabernet Sauvignon (C), Malbec (Ma), Merlot (Me), Sangiovese (Sa), Syrah (Sy) and Tempranillo (T)], harvested in 2012 from thirteen collaborating wineries of Mendoza and San Juan (Argentina), including an experimental winery from Facultad de Ciencias Agrarias (FCA), Universidad Nacional de Cuyo, Mendoza, Argentina. The thirteen wineries were: Galán (A, B, C, Ma, Me, T), CoViTu (B, C, Ma, Me, T), experimental winery FCA (C, Ma, Me), San Rafael (Ma, Sy), Agrelo (Ma, Me, Sa), San Juan (Cs, Ma -two samples-, Sy), Mayor Drummond (Cs), La Consulta (Sy), Plantago (Ma), and Albahaca (Ma). The wine samples from each winery were collected directly from fermentation tanks at the end of malolactic fermentation, transferred under nitrogen to completely filled amber glass bottles, and stored at 4 °C to ensure their preservation until their analysis in the laboratory.

2.3. HPLC-DAD

The optimization of HPLC method was based on the work developed by de Villiers *et al.* [8]. Prior to analysis, wine samples were filtered through a 0.45 µm pore size nylon membrane (Aura Industries Inc., New York, USA) without further treatment, and 20 µL of every sample were injected directly into the chromatographic system, consisting of a Hewlett-Packard 1100 series HPLC equipped with a degasser model G1322A, a quaternary pump model G1311A, and a photodiode array detector model G1315A (Agilent Technologies, Palo

Alto, USA). Separation was performed on a reversed-phase column Lichrocart 250-4 Purospher STAR RP-18e column (Merck, Argentina) (250 mm × 4 mm, 5 µm particle size) with a Security Guard Gemini C18 guard cartridge (Phenomenex, USA) (4 mm × 3 mm) at 25 °C. Two mobile phases were employed for elution: A (water/formic acid, 99:1, v/v) and B (acetonitrile/formic acid, 99:1, v/v), and the gradient profile was as follows: 0% B (min 0); 3% B (min 1); 15% B (min 10); 30% B (min 25); 50% B (min 35); 95% B (min 40); and 0% B (min 45). The flow rate was 1.0 mL min^{-1}. Each sample was run by triplicate, and good repeatability was observed. No changes were detected in cromatographic parameters as retention time, and peak shapes and areas in a reference sample that was run at the beginning and at the end of the analysis. All the analyses were conducted with the same guard column cartridge, keeping the maximum working pressure in the range 165-170 bar, being 250 bar the maximum recommended working pressure for the column used in this study. Diode array detection proceeded from 200 to 600 nm with a bandwidth of 2 nm and a data acquisition of five points per second. The presence of formic acid in the elution solvents is needed to maintain the pH below 2.5, thus ensuring that anthocyanins are present as a single species (flavylium cation).

2.4. Software

All calculations were made using MATLAB software (version 7.0, The Mathworks Inc., USA). Chromatographic time alignment was performed using the COSHIFT algorithm [30] included in the software developed by Tomasi *et al.* [31]. MCR-ALS was implemented using the graphical interface provided by Tauler in his web page http://www.mcrals.info/ [32]. Principal component analysis was run using an in-house MATLAB code. All programs were run on a HP Pavilion dv5-2043la microcomputer with an Intel Pentium P6000, 1.86 GHz microprocessor and 6 GB of RAM. UV-Visible data were exported from the HPLC-DAD

system as .csv (comma separated values) using the HP ChemStationRev.A.05.02 software for subsequent data processing under MATLAB.

Preliminary LC-DAD data analysis showed absorption signals in the range 200 to 260 nm corresponding to the HPLC solvent that were subtracted from the original data before chemometric analysis. To carry out this study in acceptable computational times, it was necessary to reduce the data obtained in the HPLC-DAD runs. Therefore, each sample subject to analysis consisted of an array of 2400×170 data points (0-40 min taken in steps of 1 s and 262-600 nm taken in steps of 2 nm, respectively).

3. Theory

3.1. MCR-ALS

The first step in MCR-ALS is to roughly estimate the number of components, which can be simply performed by visual inspection of singular values or principal component analysis (PCA) plots for the experimental data matrix [32,33]. This initial number of components can be afterwards refined, checking for their fit and reliability. The assumed bilinear model in MCR-ALS is analogous to the generalized Lambert-Beer's law, where the individual responses of each component are additive. In matrix form, this model is expressed as:

$$\mathbf{D} = \mathbf{C}\,\mathbf{S}^{T} + \mathbf{E} \quad (1)$$

where **D** (size $J\times K$) is the matrix of experimental data, **C** (size $J\times N$) is a matrix whose columns contain the concentration profiles of the N components present in the samples, $\mathbf{S}^{T}$ (size $N\times K$) is a matrix whose rows contain the component spectra and **E** (size $J\times K$) collects the experimental error and the variance not explained by the bilinear model of equation (1).

The resolution is accomplished using an iterative ALS procedure [33-35]. In each iteration, new **C** and $\mathbf{S}^{T}$ matrices are obtained under a series of constraints (non-negativity,

unimodality, closure, etc.) to give physical meaning to the solutions, to limit their possible number for the same data fitting, and to decrease the extent of possible rotation ambiguities [36]. Iterations continue until an optimal solution is obtained that fulfils the postulated constraints and the established convergence criterion.

The procedure described above can be easily extended to the simultaneous analysis of multiple data sets or data matrices if they have at least one data mode (direction) in common. For instance, if the different data sets have been analyzed by the same spectroscopic method, the possible data arrangement and bilinear model extension is given by the following equation:

$$\mathbf{D}_{\text{aug}} = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \cdots \\ \mathbf{D}_I \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \cdots \\ \mathbf{C}_I \end{bmatrix} \mathbf{S}^{\mathrm{T}} + \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \cdots \\ \mathbf{E}_I \end{bmatrix} = \mathbf{C}_{\text{aug}} \mathbf{S}^{\mathrm{T}} + \mathbf{E}_{\text{aug}} \qquad (2)$$

where $\mathbf{D}_{\text{aug}}$ is the augmented data matrix, constructed from I individual data matrices [37]: $\mathbf{D}_1$, $\mathbf{D}_2$, ...$\mathbf{D}_I$. Each of these data matrices has size $J \times K$, where J is the number of rows and K is the number of columns. In this column-wise augmentation mode, the data matrices are placed on top of each other, giving the matrix $\mathbf{D}_{\text{aug}}$ of size $IJ \times K$, which keeps the same number of columns in all of them, and where the different data matrices share their column vector space, $\mathbf{C}_{\text{aug}}$ is the column-wise augmented matrix of size $IJ \times N$, and $\mathbf{E}_{\text{aug}}$ is the corresponding augmented error matrix.

In the case of data matrices augmented row-wise, the individual data matrices are placed one adjacent to the other, giving the matrix $\mathbf{D}_{\text{aug}}$ of size $J \times IK$, which keeps the same number of rows in all of them, and where the row vector space is shared:

$$\mathbf{D}_{\text{aug}} = [\mathbf{D}_1 \mathbf{D}_2 \ \ldots \ \mathbf{D}_I] = \mathbf{C} \left[\mathbf{S}_1^{\mathrm{T}} \quad \mathbf{S}_2^{\mathrm{T}} \quad \ldots \quad \mathbf{S}_I^{\mathrm{T}} \right] + \left[\mathbf{E}_1^{\mathrm{T}} \quad \mathbf{E}_2^{\mathrm{T}} \quad \ldots \quad \mathbf{E}_I^{\mathrm{T}} \right] =$$

$$= \mathbf{C}\, \mathbf{S}_{\text{aug}}^{\mathrm{T}} + \mathbf{E}_{\text{aug}}^{\mathrm{T}} \qquad (3)$$

where all symbols are as in equation (2). When data fulfill the trilinear model, both type of matrix augmentations, column- and row-wise, are equivalent. However, when data do not fulfill the trilinear model (but they still fulfill the bilinear model), the two types of augmentation are not equivalent: matrix augmentation should be performed in the mode where chemical rank (mathematical rank in absence of noise) is better preserved, i.e., where it is equal to the number of chemical constituents. This implies that the response profiles of the components in this mode are invariant, and do not change from sample to sample. In many cases, particularly in chromatographic-spectral analysis, such a situation is not achieved, and the chemical rank is only preserved in one of the two modes of matrix augmentation [37]. In this latter case it is usual to perform a column-wise augmentation sharing the spectral subspace among the samples, because of experimental changes in elution profiles from run to run, both in shape and peak position. However, this requires that the various sample component present different spectra, so that selectivity is achieved in the spectral mode. Column-wise augmentation was initially attempted in this work, but several sample components showed almost identical spectra (e.g., all anthocyanin compounds absorbing at ca. 520 nm cannot be resolved from each other in this way). Therefore, we decided to employ the less common augmentation by sharing the time subspace, or row-wise augmentation above [27,28]. However, this requires that the elution time traces were aligned before MCR-ALS data processing, in order to have a common elution profile for a given component in all samples. After decomposition in this augmentation mode, the scores for each constituent are computed as the sum of the elements of the corresponding profile in each of the sub-matrices of $\mathbf{S}_{aug}$ according to:

$$a_{i,n} = \sum_{k=1}^{K} s_i(k,n) \qquad (4)$$

where *i* identifies the sample, *n* the constituent, *j* each of the data points or channels in the sub-matrix along the non-augmented mode and $s_i(k,n)$ the element of the $\mathbf{S}_i$ matrix [see equation (3)] at channel *k* for component *n*.

3.2. Principal component analysis

After MCR-ALS decomposition of the augmented matrix, a matrix of scores is obtained, of size $I \times N$ (I = number of samples, N = number of constituents), which could in principle be employed for sample discrimination. However, if $N > 3$ it is preferable to reduce the dimensionality of the score matrix using PCA, which usually concentrates the variance in a smaller number of principal components (PCs). Usually two of them are employed to build a plot of sample positions in score space, achieving sample discrimination. The outcome of PCA is thus: (1) the PC values for each sample, from which the first two are used for discrimination, and (2) a loading matrix, which shows the relative contribution of each MCR-ALS score to each of the PC, helping to choose the true discriminating variables [38].

4. Results and discussion

4.1. LC-DAD data pre-processing

Figure 2A shows the chromatographic-spectral landscape obtained for a specific sample (Aspiran *varietal*, Galán winery) after injection into the HPLC-DAD system. From this latter Figure, specific chromatographic traces can be obtained at selected wavelengths: Fig. 2B and 2C show the corresponding elution time profiles for the same sample at 280 and 520 nm respectively. Due to the fact that most of the flavonoids absorb at 280 nm, Fig. 2B shows a large number of unresolved components at this latter wavelength. On the other hand, comparatively less components appear in the chromatogram of Fig. 2C at 520 nm, which however implies the presence of several anthocyanin compounds.

The complexity of the studied samples, which can be gathered from the visual inspection of Fig. 2, requires suitable data processing algorithms to extract hidden features, or to resolve individual sample components in terms of their chromatograms and spectra. Among the various algorithms allowing to process sets of data matrices such as those presently studied, one should select a methodology which is able to model the particular data structure at hand. One specific property of the present data is the existence of changes in the elution time profile for a given component from run to run. The algorithm of choice under these conditions is multivariate curve resolution-alternating least-squares (MCR-ALS). As discussed in a previous section, this latter methodology frequently builds an augmented data matrix by placing all individual sample matrices adjacent to each other in a column-wise augmentation mode. This allows one to model, after suitable constraints during the fitting phase, the varying time profiles of the sample components in the various samples.

However, the successful application of this technique requires that sufficient selectivity exists in the spectral mode. If several sample components display very similar or identical spectra, they cannot be resolved into individual components by MCR-ALS. In this case, one viable alternative is to perform a row-wise matrix augmentation [27,28]. The requirements for resolution in this augmentation mode are: (1) selectivity in the chromatographic data mode, and (2) time synchronization or alignment of the chromatograms in such a way that component peaks have the same shape (although the area under the peak may differ) in different samples.

Many different algorithms are available for chromatographic-spectral matrix alignment [29]. Some of them shift an entire chromatographic matrix with respect to a reference one by a number of data points, without modifying the peak shapes or the time distance between peaks. More powerful methodologies exist, however, which are able to warp the chromatograms, changing peak positions and shapes. They are in principle necessary to

process long chromatographic runs such as those presently studied. Among the latter ones the following have been reported: Interval Correlation Optimized Shifting (ICOSHIFT) [39], Dynamic Multi-way Warping (DMW) [31], Correlation Optimized Warping (COW) [40], Correlation Optimized Shifting (COSHIFT) [31], etc. All these possibilities were probed to the present data, with optimum results using the latter COSHIFT algorithm, which operates by shifting a data matrix in both the row and column directions, in order to get maximum matrix-correlation from the RV-coefficient (which is a multivariate generalization of the squared Pearson correlation coefficient), assuming that peak widths are invariant. It is important to notice, in this regard, that we did not detect significant changes in chromatographic peak shapes from run to run. As an example, Fig. 3 shows a zoom selection of the chromatographic profile at 280 nm of a typical sample (Malbec *varietal*, Galán winery) before (blue line) and after (red line) the application of this algorithm, in which the finally obtained alignment is apparent. The sample used as reference (black line, Merlot *varietal*, CoViTu winery) was utilized as reference for the alignment of the remaining ones.

4.2. MCR-ALS resolution of LC-DAD data

After COSHIFT chromatographic alignment of all samples, MCR-ALS analysis was applied to the row-wise augmented data matrix, namely, an array of 2400 × 4590 data points as explained in Section 3.1. As a first step before data resolution, the number of components was estimated by principal component analysis of the augmented data matrix, inspecting a plot of singular values as a function of increasing number of trial components. In this way, 10 components were selected, which explained 94.37% of the data variance; after the tenth component, no further significant decrease in the singular values was detected. Additionally, this initial estimate was confirmed by processing the LC-DAD data with MCR-ALS with more components as initial estimate (i.e., 12, 15 and 20, with 95.41%, 96.54% and 97.67%

explained variance, respectively), with results which did not significantly differed from those obtained with 10 components as initial estimate. On the other hand, principal component analysis of the column-wise augmented data matrix, namely, an array of 64800 × 170 data points, as explained in Section 3.1, showed that only 3 components were needed to explain 97.39% of the data variance (components 4 and 5 only explained 1.37% and 0.52% respectively), revealing that more components can in principle be resolved in the row-wise augmented data matrix.

In order to achieve successful resolution, non-negativity in both spectra and chromatograms was applied during the least-squares fit, until successive changes in residual fit were smaller than 0.1%. This typically required 30 iterations. MCR-ALS resolution was obtained with good quality parameters, namely, fitting error (L.O.F.) of 5.99% and 7.69% (regarding PCA and experimental respectively) and 99.41% of explained variance. The result is shown in Fig. 4, in the form of a common chromatographic profile for the 10 resolved constituents (Fig. 4A) and a fragment of the augmented spectra corresponding to four selected samples (Fig. 4B). Figure 4 shows that several sample components were resolved with maxima at ca. 520 nm, corresponding to different anthocyanin compounds. This would not be possible in the usual column-wise augmentation mode, because in the latter mode a single spectrum is retrieved for all anthocyanin compounds, and our intention was to differentiate these important class of compounds from each other.

As fingerprint information, MCR-ALS renders the area under the resolved spectral profile for each component in a particular sample. This information was arranged into a matrix of size 27 × 10 (27 samples × 10 constituents). In order to reduce the dimensionality of this latter matrix for intuitive discrimination purposes, principal component analysis was applied to this fingerprint matrix, as discussed in the next section.

4.3. PCA discrimination using MCR-ALS scores

In order to study the relation among the MCR-ALS fingerprint information with the eight wine *varietals* and the thirteen wineries, the output score matrix was subjected to principal component analysis (PCA). Figure 5A shows a typical score plot of first *vs*. second principal component (45.40% and 24.06% of variance retained by PC1 and PC2 respectively). In this Figure, we can observe a partial discrimination into winery provenance (i.e., geographical origin) of the samples corresponding to Galán, CoViTu, San Juan, and San Rafael wineries from the remaining samples. Moreover, a plot of first vs. third principal component (PC3, 13.40% variance retained), shown in Fig. 5B, reveals that all samples corresponding to the Malbec *varietal* (the *insignia* argentine *varietal*) are discriminated from the remaining samples.

Examination of the contribution of the constituents resolved by MCR-ALS in each principal component reveals which compounds were decisive for wine discrimination (Fig. 6A). Constituents No. 3 and 10 displayed the largest contributions to PC1 and PC2. For PC3, on the other hand, in addition to No. 10, a contribution from No. 2 is detected. Figure 6B shows the resolved spectra of the relevant constituents, in which it can be observed that constituents No. 2 and 3 have spectra with absorption maxima at 520 nm (anthocyanins), whereas constituent No. 10 has a spectrum with an absorption maximum at 330 nm (phenolic acids). This means that different anthocyanin compounds contribute for discrimination by geographical origin, whereas for Malbec *varietal* discrimination from the rest of the samples, both anthocyanins and phenolic acids are needed.

5. Conclusions

Wine study was carried out by direct sample injection HPLC-DAD without sample pre-treatment. The obtained data were processed by MCR-ALS in the form of an augmented

data matrix, with a less common row-wise augmentation with the data matrices sharing the time subspace. To achieve this, it was necessary to perform previous time alignment of the chromatograms using the COSHIFT algorithm. The matrix of sample scores resolved by MCR-ALS was then submitted to PCA, which allowed discriminating all Malbec *varietals* from the remaining samples, and also to explore the wine samples by geographical origin, in this case with only partial success. The results here obtained are promising. Analysis of the constituents of each principal component showed that anthocyanin compounds present in wine were crucial to perform both types of discrimination.

Acknowledgements

We acknowledge financial support from Universidad Nacional de Rosario, Universidad Nacional de Cuyo, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2010-0084). P.L.P. thanks CONICET for a postdoctoral fellowship.

References

[1] R.S. Jackson, 6- Chemical constituents of grapes and wine, in: Wine science (Third edition), Academic Press, San Diego, 2008, pp. 270-331.

[2] M.V. Moreno-Arribas, M.C. Polo, 9- Phenolic compounds, in: M.V. Moreno-Arribas, M.C. Polo (Eds.) Wine chemistry and biochemistry, Springer New York, 2009, pp. 437-527.

[3] M.V. Moreno-Arribas, M.C. Polo, 9D- Influence of phenolics on wine organoleptic properties, in: M.V. Moreno-Arribas, M.C. Polo (Eds.) Wine chemistry and biochemistry, Springer New York, 2009, pp. 529-570.

[4] J. Saurina, Characterization of wines using compositional profiles and chemometrics, Trends Anal. Chem., 29 (2010) 234-245.

[5] A. Tredoux, A. de Villiers, P. Májek, F.d.r. Lynen, A. Crouch, P. Sandra, Stir bar sorptive extraction combined with GC-MS analysis and chemometric methods for the classification of South African wines according to the volatile composition, J. Agric. Food Chem., 56 (2008) 4286-4296.

[6] D. Ballabio, T. Skov, R. Leardi, R. Bro, Classification of GC-MS measurements of wines by combining data dimension reduction and variable selection techniques, J. Chemom., 22 (2008) 457-463.

[7] R.F. Alves, A.M.D. Nascimento, J.M.F. Nogueira, Characterization of the aroma profile of Madeira wine by sorptive extraction techniques, Anal. Chim. Acta, 546 (2005) 11-21.

[8] A.d. Villiers, G. Vanhoenacker, P. Majek, P. Sandra, Determination of anthocyanins in wine by direct injection liquid chromatography–diode array detection–mass spectrometry and classification of wines using discriminant analysis, J. Chromatogr., A, 1054 (2004) 195-204.

[9] M. Fanzone, Á. Peña-Neira, V. Jofré, M. Assof, F. Zamora, Phenolic characterization of Malbec wines from mendoza province (Argentina). J. Agric. Food Chem., 58 (2010) 2388-2397.

[10] L. Vaclavik, O. Lacina, J. Hajslova, J. Zweigenbaum, The use of high performance liquid chromatography-quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety, Anal. Chim. Acta, 685 (2011) 45-51.

[11] A. Cuadros-Inostroza, P. Giavalisco, J. Hummel, A. Eckardt, L. Willmitzer, H. Peña-Cortés, Discrimination of wine attributes by metabolome analysis, Anal. Chem., 82 (2010) 3573-3580.

[12] L. Jaitz, K. Siegl, R. Eder, G. Rak, L. Abranko, G. Koellensperger, S. Hann, LC–MS/MS analysis of phenols for classification of red wine according to geographic origin, grape variety and vintage, Food Chem., 122 (2010) 366-372.

[13] M. Anastasiadi, A. Zira, P. Magiatis, S.A. Haroutounian, A.L. Skaltsounis, E. Mikros, ^{1}H NMR-based metabonomics for the classification of Greek wines according to variety, region, and vintage. Comparison with HPLC data, J. Agric. Food Chem., 57 (2009) 11067-11074.

[14] J.-E. Lee, G.-S. Hwang, F. Van Den Berg, C.-H. Lee, Y.-S. Hong, Evidence of vintage effects on grape wines using ^{1}H NMR-based metabolomic study, Anal. Chim. Acta, 648 (2009) 71-76.

[15] D. Cozzolino, H.E. Smyth, M. Gishen, Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins, J. Agric. Food Chem., 51 (2003) 7703-7708.

[16] L. Liu, D. Cozzolino, W.U. Cynkar, R.G. Dambergs, L. Janik, B.K. O'Neill, C.B. Colby, M. Gishen, Preliminary study on the application of visible–near infrared spectroscopy and

chemometrics to classify Riesling wines from different countries, Food Chem., 106 (2008) 781-786.

[17] R. Garrido-Delgado, S. López-Vidal, L. Arce, M. Valcárcel, Differentiation and identification of white wine varieties by using electropherogram fingerprints obtained with CE, J. Sep. Sci., 32 (2009) 3809-3816.

[18] J. Pazourek, D. Gajdošová, M. Spanilá, M. Farková, K. Novotná, J. Havel, Analysis of polyphenols in wines: Correlation between total polyphenolic content and antioxidant potential from photometric measurements: Prediction of cultivars and vintage from capillary zone electrophoresis fingerprints using artificial neural network, J. Chromatogr., A, 1081 (2005) 48-54.

[19] M.P. Fabani, R.C. Arrúa, F. Vázquez, M.P. Diaz, M.V. Baroni, D.A. Wunderlin, Evaluation of elemental profile coupled to chemometrics to assess the geographical origin of Argentinean wines, Food Chem., 119 (2010) 372-379.

[20] P. Pohl, What do metals tell us about wine?, Trends Anal. Chem., 26 (2007) 941-949.

[21] D.P. Makris, S. Kallithraka, A. Mamalos, Differentiation of young red wines based on cultivar and geographical origin with application of chemometrics of principal polyphenolic constituents., Talanta, 70 (2006) 1143-1152.

[22] N.H. Beltrán, M.a. Duarte-Mermoud, M.a. Bustos, S.a. Salah, E.a. Loyola, A. Peña-Neira, J.W. Jalocha, Feature extraction and classification of Chilean wines, J. Food Eng., 75 (2006) 1-10.

[23] S.a. Bellomarino, X.a. Conlan, R.M. Parker, N.W. Barnett, M.J. Adams, Geographical classification of some Australian wines by discriminant analysis using HPLC with UV and chemiluminescence detection., Talanta, 80 (2009) 833-838.

[24] M. Fanzone, Á. Peña-Neira, M. Gil, V. Jofré, M. Assof, F. Zamora, Impact of phenolic and polysaccharidic composition on commercial value of Argentinean Malbec and Cabernet Sauvignon wines, Food Res. Int., 45 (2012) 402-414.

[25] D. Serrano-Lourido, J. Saurina, S. Hernández-Cassou, A. Checa, Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis, Food Chem., 135 (2012) 1425-1431.

[26] E. Salvatore, M. Cocchi, A. Marchetti, F. Marini, A. de Juan, Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC-DAD and chemometric techniques, Anal. Chim. Acta, 761 (2013) 34-45.

[27] A. Mancha de Llanos, M.M. Zan, M.J. Culzoni, A. Espinosa-Mansilla, F. Cañada-Cañada, A.M. Peña, H.C. Goicoechea, Determination of marker pteridines in urine by HPLC with fluorimetric detection and second-order multivariate calibration using MCR-ALS, Anal. Bioanal. Chem., 399 (2011) 2123-2135.

[28] M.J. Culzoni, A. Mancha de Llanos, M.M. De Zan, A. Espinosa-Mansilla, F. Cañada-Cañada, A. Muñoz de la Peña, H.C. Goicoechea, Enhanced MCR-ALS modeling of HPLC with fast scan fluorimetric detection second-order data for quantitation of metabolic disorder marker pteridines in urine, Talanta, 85 (2011) 2368-2374.

[29] T.G. Bloemberg, J. Gerretzen, A. Lunshof, R. Wehrens, L.M.C. Buydens, Warping methods for spectroscopic and chromatographic signal alignment: A tutorial, Anal. Chim. Acta, 781 (2013) 14-32.

[30] V.G. van Mispelaar, A.C. Tas, A.K. Smilde, P.J. Schoenmakers, A.C. van Asten, Quantitative analysis of target components by comprehensive two-dimensional gas chromatography, J. Chromatogr., A, 1019 (2003) 15-29.

[31] G. Tomasi, F. van den Berg, C. Andersson, Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data, J. Chemom., 18 (2004) 231-241.

[32] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB, Chemom. Intell. Lab. Syst., 76 (2005) 101-110.

[33] M. Maeder, A. Zilian, Evolving factor analysis, a new multivariate technique in chromatography, Chemom. Intell. Lab. Syst., 3 (1988) 205-213.

[34] M. Maeder, Evolving factor analysis for the resolution of overlapping chromatographic peaks, Anal. Chem., 59 (1987) 527-530.

[35] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem., 63 (1991) 1425-1432.

[36] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom., 9 (1995) 31-58.

[37] R. Tauler, M. Maeder, A. de Juan, 2.24- Multiset data analysis: Extended multivariate curve resolution, in: Editors-in-Chief:, D.B. Stephen, T. Romà, W. Beata (Eds.) Comprehensive chemometrics, Elsevier, Oxford, 2009, pp. 473-505.

[38] I.T. Jolliffe, Principal component analysis, 2nd ed., Springer, New York, 2002.

[39] G. Tomasi, F. Savorani, S.B. Engelsen, *i*coshift: An effective tool for the alignment of chromatographic data, J. Chromatogr., A, 1218 (2011) 7832-7840.

[40] N.-P.V. Nielsen, J.M. Carstensen, J. Smedsgaard, Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping, J. Chromatogr., A, 805 (1998) 17-35.

Figure captions

Fig. 1. Representative structures of the three main families of phenolic compounds found in wine.

Fig. 2. A) A typical chromatographic-wavelength landscape. B) Chromatographic trace at 280 nm. C) Chromatographic trace at 520 nm.

Fig. 3. Illustration of the application of the COSHIFT algorithm for chromatographic alignment to a typical sample. Black line, reference trace at 280 nm, blue line, an unaligned chromatogram at the same wavelength, and red line, aligned chromatogram.

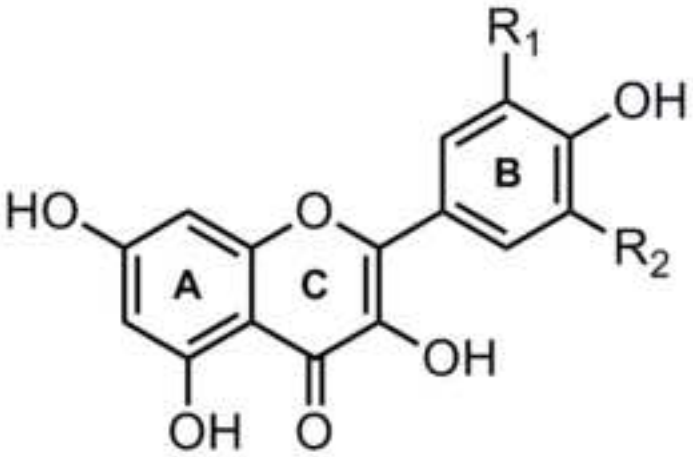
Fig. 4. Profiles for the ten constituents resolved by MCR-ALS from the augmented data matrix in the spectral direction. A) Elution time profiles. B) Augmented spectral profiles (only four representative samples are shown).

Fig. 5. Discrimination of wine samples from principal component analysis. A) PC2 vs. PC1. B) PC3 vs. PC1.

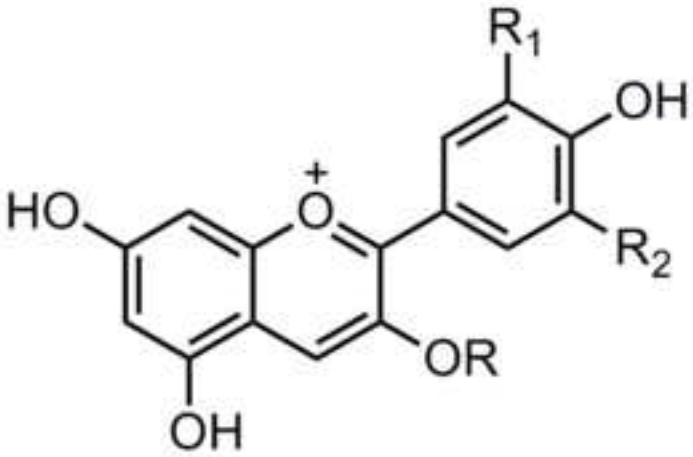
Fig. 6. A) Loading composition of the first three principal components, in terms of the ten MCR-ALS resolved components. B) MCR-ALS resolved spectra of components No. 2, 3 and 10.

Figure 1

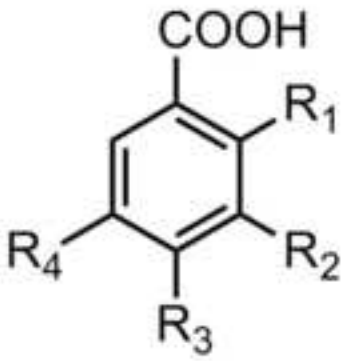
Click here to download high resolution image



flavonols



R= H, anthocyanidins
R= carbohydrate, anthocyanins



phenolic acids

Figure 2

Click here to download high resolution image

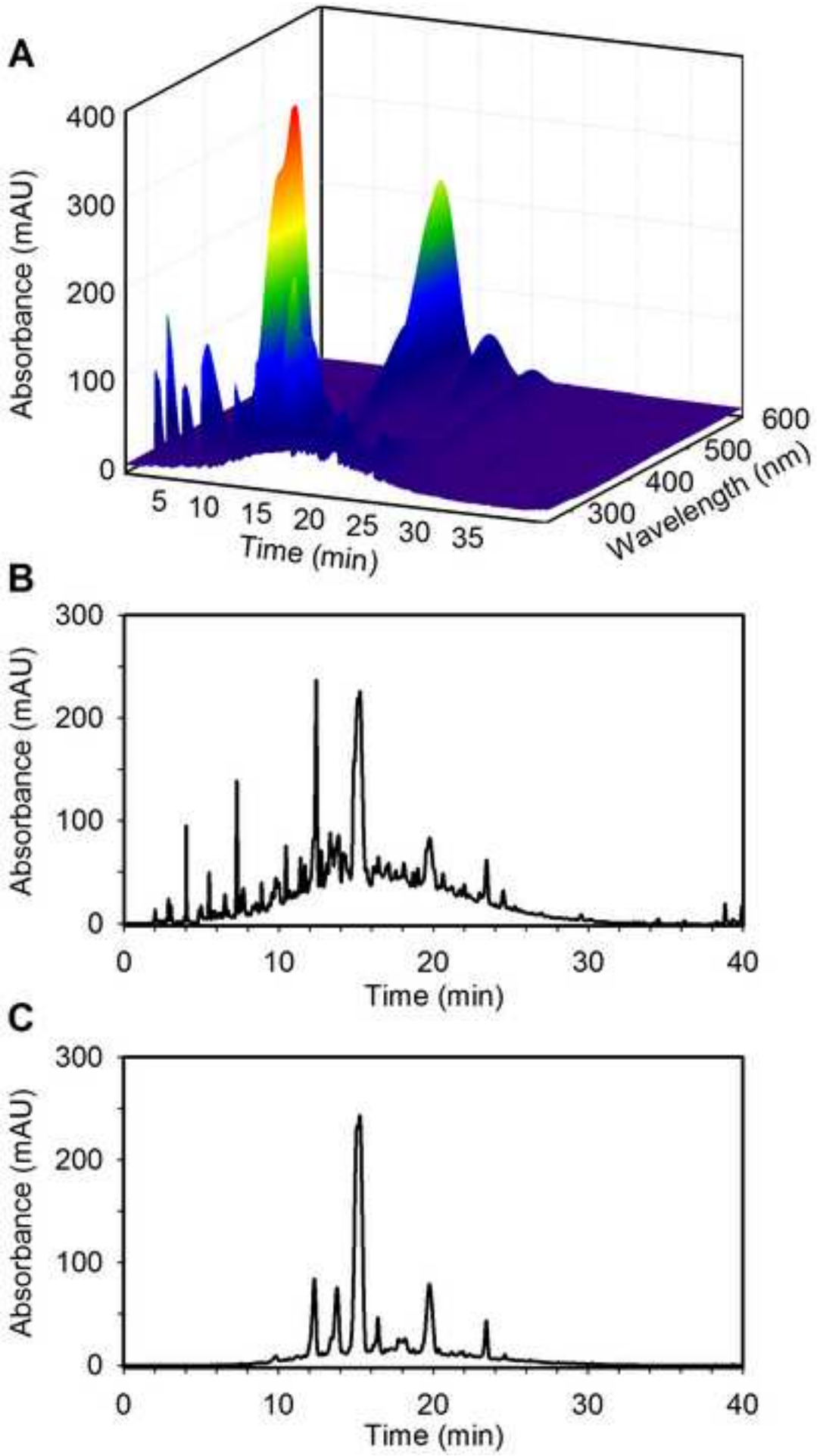


Figure 3

Click here to download high resolution image

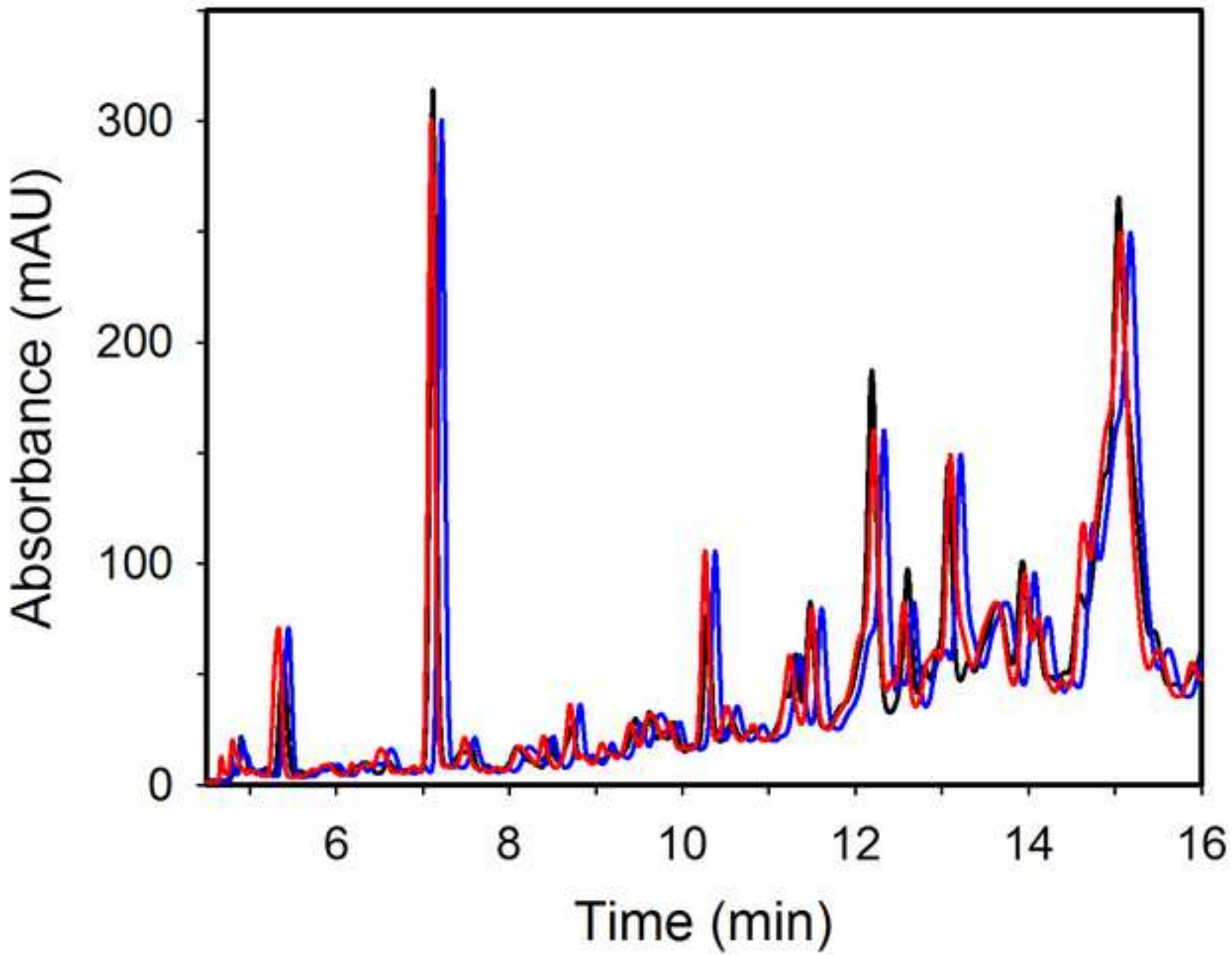


Figure 4
Click here to download high resolution image

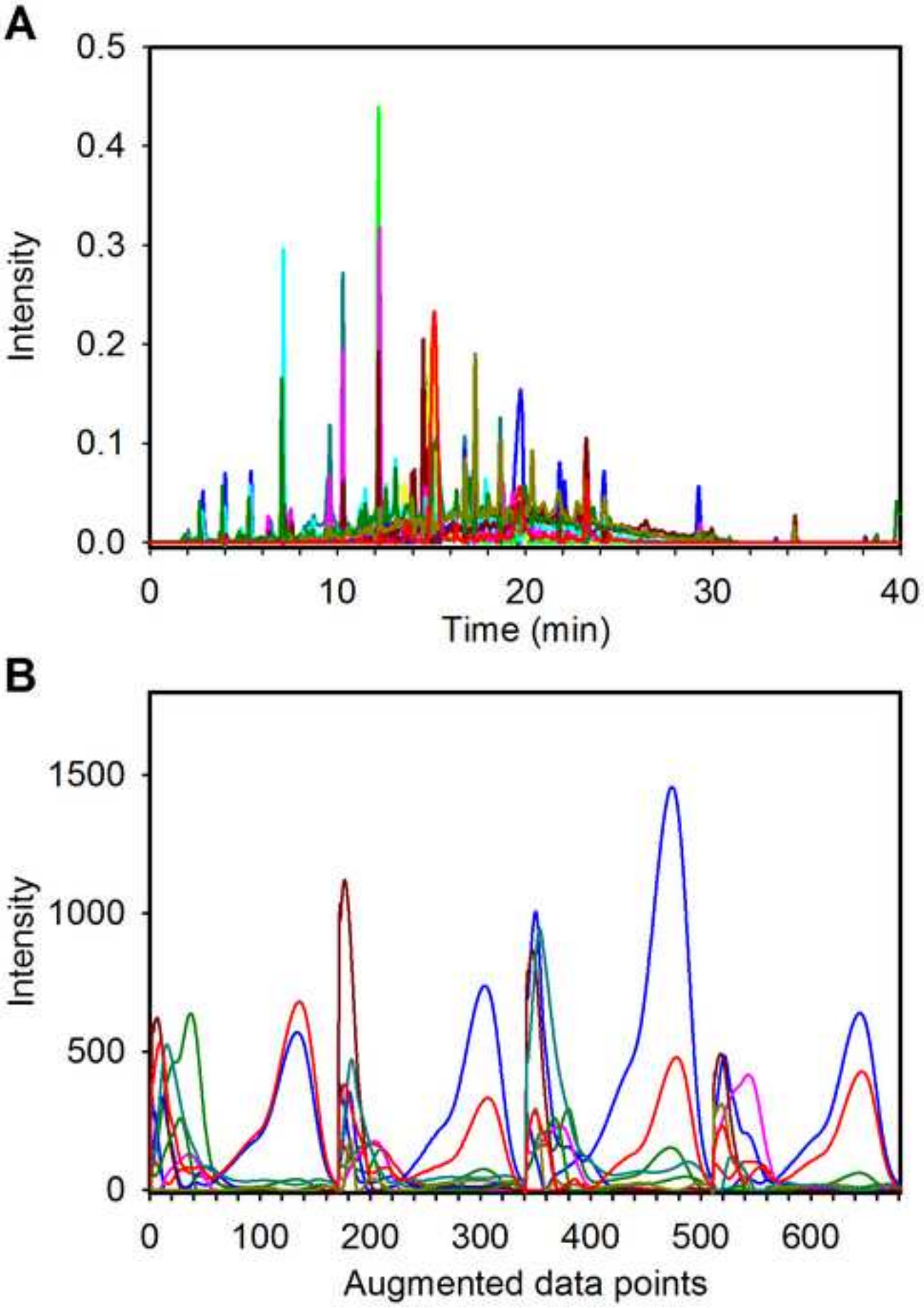


Figure 5

Click here to download high resolution image

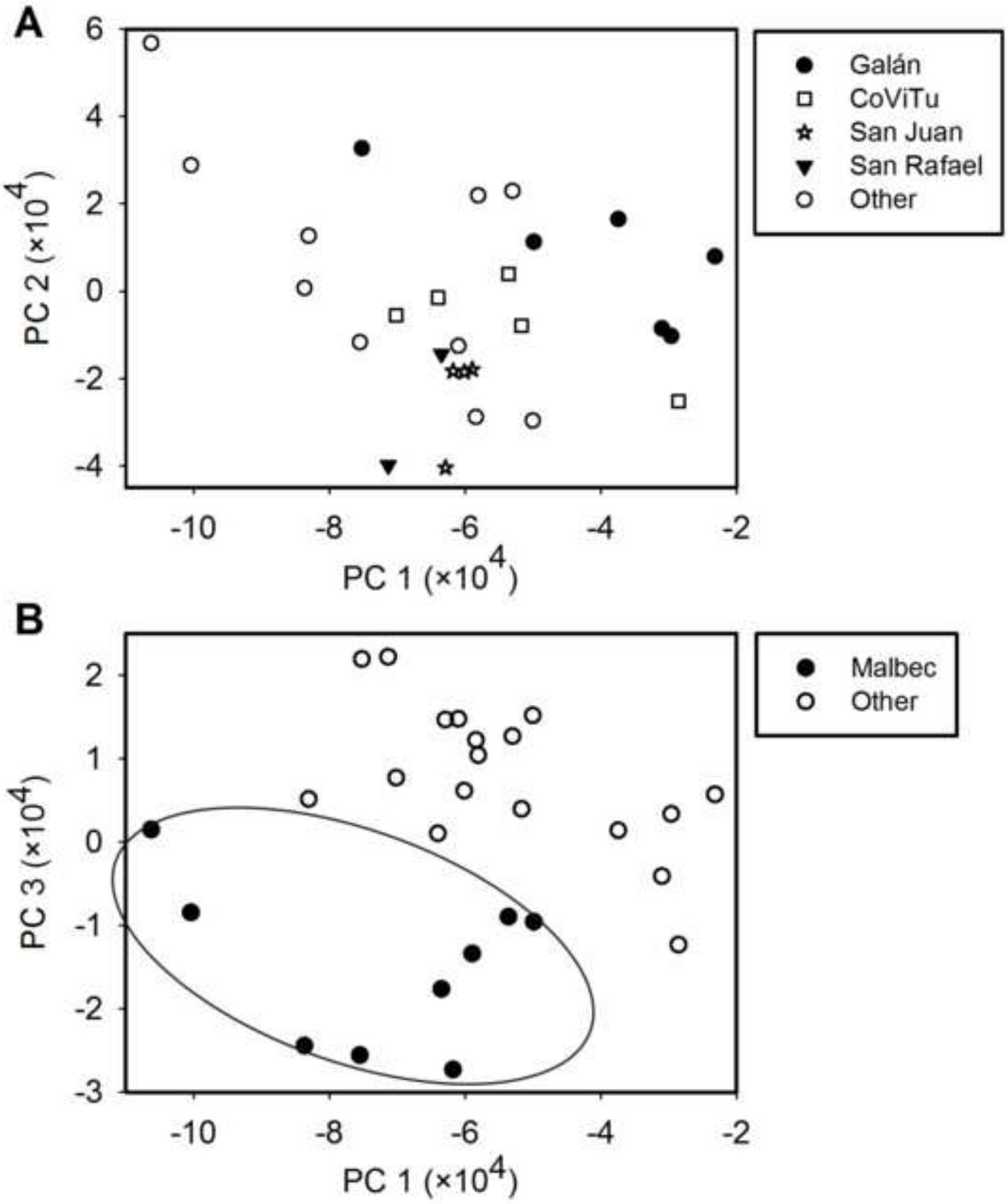


Figure 6
Click here to download high resolution image