Talanta 88 (2012) 303–310

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Feasibility study of FT-MIR spectroscopy and PLS-R for the fast determination of anthocyanins in wine

M. Romera-Fernández[a,1], L.A. Berrueta[a,1], S. Garmón-Lobato[a,1], B. Gallo[a,*], F. Vicente[a,1], J.M. Moreda[b]

[a] Department of Analytical Chemistry, Faculty of Science and Technology, Universidad del País Vasco/Euskal Herriko Unibertsitatea, P.O. Box 644, 48080 Bilbao, Spain
[b] R+D Department, Grupo Faustino, Crta. Logroño s/n, 01320 Oyón (Álava), Spain

ARTICLE INFO

Article history:
Received 21 July 2011
Received in revised form 5 October 2011
Accepted 16 October 2011
Available online 31 October 2011

Keywords:
Wine analysis
Anthocyanins
FT-IR
HPLC–DAD
PLS regression

ABSTRACT

The feasibility of using Fourier Transform Mid-Infrared Spectroscopy (FT-MIR) combined with Partial Least Squares Regression (PLS-R) for the determination of 12 anthocyanins (3-*O*-glucosides of delphinidin, cyanidin, petunidin, peonidin and malvidin, as well as acetic acid esters and *p*-coumaric acid esters of petunidin, peonidin and malvidin and caffeic acid ester of malvidin) and three sums (sum of non-acylated anthocyanins, sum of acetylated anthocyanins and sum of coumaroylated anthocyanins), in red wines has been tested. Reference values of anthocyanin concentrations by reverse-phase High Performance Liquid Chromatography with Diode Array Detection (HPLC–DAD) were used to calibrate the models. A Principal Component Analysis (PCA) was applied to these reference values and a differentiation of wine samples by wine type (*young* wines of 2005, *young* wines of 2004 and *crianza* and *reserva* wines) has been possible.

A calibration model using PLS-R was built with 153 samples of *Rioja* wines and the prediction of the anthocyanin concentrations using this model was evaluated by internal and external validation sample sets. Most of the anthocyanins and their sums have been predicted with a Standard Error of Prediction (SEP) of 15–30% for *young* wines recently bottled. However, for *young* wines after one year of being bottled, and for *crianza* and *reserva* wines, these errors were unacceptable. The obtained results suggest that the model built for FT-IR instrument calibration is a useful tool for a quick determination of the anthocyanin content of *young* wines of the current vintage, but a careful robust external validated calibration of the technique is necessary in order to maintain the prediction errors within controlled limits.

1. Introduction

Currently, infrared spectroscopy is one of the most common spectroscopic techniques used by the industry. It has been employed with different aims like identification and quantification of interesting compounds in food (meat [1,2], fish [3], milk and dairy products [4,5], cereals [6,7], fruit [8], vegetables [9], fruit juices [10], wine [11], beer [12], eggs and derived products from them [13]), detection of adulteration and contamination, characterization, authentication or classification.

The use of NIR has the longest history [14–17], but nowadays the application of FT-MIR is especially interesting for wine analysis [16,18–23]. Nowadays, the exhaustive control that wineries subject their products, with the purpose of obtaining wines of greater quality, has motivated that FT-MIR combined with advances in chemometrics (multivariate data analysis) have been incorporated in the enological field. A large number of chemical parameters of great importance for the wine quality are determined by FT-MIR [24–31], such as ethanol, glycerol, alcoholic degree, total acidity, pH, SO_2, reducing sugars, glucose, fructose, volatile acidity, tartaric acid, gluconic acid, lactic acid, malic acid, acetic acid, citric acid and Folin Ciocalteu index. . . All these parameters can be measured in a few minutes in only one analysis with little previous sample preparation and variable error levels (RSD < 5.0% for major components such as ethanol or total organic acids, but about 20% for minor components such as sugars). Therefore, this method for quality assurance purposes is widely used in the wine industry as a substitute to conventional chemical methods of wine analysis involving time-consuming and laborious procedures [32,33]. However, some disadvantages should be taken into consideration, such as the relative high investment cost, limitation on measuring low concentration compounds, possible influence of unknown sample adulteration on the performance of calibration model, higher prediction errors than classical chemical methods, need of robust calibrations that require additional cost of time, samples and

* Corresponding author. Tel.: +34 94 601 53 90; fax: +34 94 601 35 00.
E-mail address: blanca.gallo@ehu.es (B. Gallo).
[1] Fax: +34 94 601 35 00.

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.10.045

complementary reference analysis to attain adequate results, or kind of wine (cultivars, vintage, production region) dependant calibration [29].

On the other hand, the determination of anthocyanins has become an important object of enology, because these polyphenolic compounds have a decisive influence on an organoleptic aspect that determines the quality of a wine: the color. Moreover, anthocyanins have beneficial effects upon human health such as antioxidant and anti-inflammatory properties, important to prevent diseases [34,35]. The first applications of FT-MIR to the quantification of anthocyanins were carried out by Versari et al. [36] for total anthocyanins and by Soriano et al. [22] for individual anthocyanins. The last one studied *young* red wines from different grape cultivars (Cencibel, Cabernet Sauvignon, Garnachatintorera, Syrah, Merlot, Bobal and Monastrell) and from six Protected Designations of Origin (La Mancha, Manchuela, Utiel-Requena, Almansa, Jumilla and Alicante).

The concept of calibration, which is widely used in analytical chemistry, is also applicable for FT-IR spectroscopy and, in order to predict the concentration of the components of interest, a previous calibration for these components is required, so an additional reference analysis of wine calibration samples is necessary. Reference analysis may be more accurate but it is also more time consuming (requiring sample preparation and even chemical manipulations) and too expensive to use routinely. On the other hand, due to the complexity of the information contained in the FT-IR spectra, an extensive calibration process that involves multivariate statistical procedures is required [37–39].

There are two main problems associated with wine anthocyanin characterization by FT-MIR spectroscopy and the application of multivariate data analysis. First, the similarity between the IR absorption characteristics of the different anthocyanins due to they are chemically very similar; furthermore, absorbance at a given wavelength may have contributions from more than one compound. Thus, the dominating absorption of ethanol, water and in some cases sugars strongly influences the determination of other components [29].

Both limitations are critical for the analysis of phenolic compounds, because ethanol, water and organic acids absorb in the same MIR region, masking the characteristic IR vibrations of phenols [40]. To overcome these problems, chemometric techniques, such as Principal Component Analysis (PCA) and Partial Least Squares Regression (PLS-R), are commonly used tools to develop mathematical models for these and other parameters of wine. Moreover, it is necessary to select the relevant spectral frequencies first and before using PLS-R. For covering a wide range of concentrations to guarantee the reliability of prediction, it is necessary to include a large number of samples in the calibration set.

The aim of the present work was to determine the feasibility of using FT-MIR combined with chemometrics for the determination of 12 anthocyanins (3-*O*-glucosides of delphinidin, cyanidin, petunidin, peonidin and malvidin, as well as acetic acid esters and *p*-coumaric acid esters of petunidin, peonidin and malvidin and caffeic acid ester of malvidin) and three sums (sum of non-acylated anthocyanins, sum of acetylated anthocyanins and sum of coumaroylated anthocyanins) in red wines of different degrees of ageing using a sample set that included wines from the Protected Designation of Origin *Rioja*. These wines have been elaborated using mainly Tempranillo as grape cultivar. A reference analysis by reverse-phase High Performance Liquid Chromatography with Diode Array Detection (HPLC–DAD) was used to calibrate the models. The prediction of the anthocyanin concentrations using the built model was also evaluated using internal and external validation sample sets.

2. Materials and methods

2.1. Wine samples

The calibration set consisted of 158 red wines; most of them belong to the Protected Designation of Origin *Rioja* (Spain) and obtained from 11 wineries. Some samples from other Spanish regions and/or other grape cultivars different from Tempranillo were included to check the possibility of extending the calibration to another region or different cultivar. The external validation set ($n = 24$ samples) was composed by 12 samples that were provided by the collaborating wineries and 12 samples purchased in the market (from different wineries from the eleven collaborating ones). The analyzed samples corresponded to red wines of different degrees of ageing: *young* (not aged in barrels) of vintages 2005 (thus bottled near to the date of analysis) and 2004 (bottled one year before), *crianza* (at least 12 months of ageing in oak barrels and 12 months in bottle), *reserva* (at least 12 months of ageing in oak barrels and 24 months in bottle) and *gran reserva* (at least 24 months of ageing in oak barrels and 36 months in bottle) wines. Table 1 presents the main characteristics of samples within the calibration and the external validation sets.

2.2. Reference method: HPLC–DAD

HPLC–DAD analysis was performed in a Hewlett–Packard (Palo Alto, CA, USA) series 1100 chromatography system equipped with a diode array detector. A reverse-phase Waters (Milford, MA) Nova-Pack C18 column (300 mm × 3.9 mm I.D., 4 μm) at 30 °C, protected with a guard column Waters Nova-Pack C18 (10 mm × 3.9 mm I.D., 4 μm) was used.

The HPLC–DAD conditions used in the analysis of wine samples, had been previously optimized and employed with satisfactory results in our laboratory [41] and were slightly modified for this study. The solvents used were: (A) an aqueous solution (0.4%, v/v) of phosphoric acid and (B) 100% HPLC-gradient quality acetonitrile, establishing the following gradient program: linear from 10 to 15% B, 0–30 min; 15% B isocratic, 30–35 min; linear from 15 to 30% B, 35–50 min; 30% B isocratic, 50–60 min; washing with 100% B and re-equilibration of the column.

The flow-rate was 0.8 mL/min and the injection volume was 50 μL. The sample vials in the automatic injector were thermostatted at 4 °C to guarantee the conservation of the samples. UV–vis spectra were recorded from 250 to 600 nm and the quantification was carried out at 530 nm. Previously to be injected into the HPLC instrument the wine samples were filtered through a 0.45 μm PTFE membrane (Pall Acrodisc CR 13; Port Washington, NY) and dark vials were used to avoid the light degradation of the anthocyanins. Quantification of all anthocyanins was performed against the same external standard and expressed as mg/L equivalents of Malvidin-3-*O*-glucoside (Extrasynthèse, Genay, France). Precision calculated as %RSD of reference values was lower than 5% and limit of quantitation was lower than 0.1 mg/L for all the anthocyanins.

The anthocyanins were identified by their relative retention times using Malvidin-3-*O*-glucoside and Malvidin-3-*O*-(6-*p*-coumaroyl)-glucoside as reference peaks, and by their UV–vis absorption spectra. In addition, these identifications have been confirmed using a Micromass (Milford, MA, USA) Quattro micro triple quadrupole mass spectrometer coupled to the exit of the diode array detector and equipped with a Z-spray ESI source (HPLC–DAD–MS/MS) (m/z values for M^+ ion and characteristic fragments are: Delphindin-3-*O*-glucoside: 465, 303, 153, 149, 121; Cyanidin-3-*O*-glucoside: 449, 287, 137, 149, 121; Petunidin-3-*O*-glucoside: 479, 317, 150; Peonidin-3-*O*-glucoside: 463, 301, 150; Malvidin-3-*O*-glucoside: 493, 331, 150; Petunidin-3-(6-*O*-acetyl)-glucoside: 521, 317, 150; Peonidin-3-(6-*O*-acetyl)-glucoside:

Table 1
Degrees of ageing, cultivars and vintages of samples within the calibration and external validation sets.

Calibration sample set	
Young wines (vintage 2005)	
53	100% Tempranillo
12	Tempranillo (major) and other cultivars
7	Other cultivars (100% Graciano, Merlot, Syrah, Garnacha, Cabernet Sauvignon)
Young wines (vintage 2004)	
8	100% Tempranillo
9	Tempranillo (major) and other cultivars
Crianza (vintages 2001–2004)	
27	100% Tempranillo
2	Tempranillo (major) and other cultivars
Reserva (vintages 1997–2003)	
31	100% Tempranillo
1	Tempranillo (major) and other cultivars
3	Other cultivars (100% Graciano, Garnacha)
Gran reserva (vintages 1995–1996)	
5	100% Tempranillo
Total: 158 samples	

External validation sample set			
Samples purchased in the market		Provided by collaborating wineries	
Young wines (vintage 2005)		*Young* wines (vintage 2005)	
2	100% Tempranillo	6	100% Tempranillo
2	Unknown		
Young wines (vintage 2004)			
1	100% Tempranillo		
1	Unknown		
Crianza (vintages 2002–2003)		*Crianza* (vintages 2002–2003)	
2	100% Tempranillo	3	100% Tempranillo
1	Tempranillo (major) and other cultivars	1	Tempranillo and other cultivars
1	Unknown		
Reserva (vintages 1998, 2000)		*Reserva* (vintages 1999, 2000)	
1	100% Tempranillo	2	100% Tempranillo
1	Unknown		
Total: 24 samples			

505, 301, 150; Malvidin-3-(6-*O*-acetyl)-glucoside: 535, 331, 150; Petunidin-3-(6-*O*-*p*-coumaroyl)-glucoside: 625, 317, 150; Peonidin-3-(6-*O*-*p*-coumaroyl)-glucoside: 609, 301, 150; Malvidin-3-(6-*O*-*p*-coumaroyl)-glucoside: 639, 331, 150; and Malvidin-3-(6-*O*-caffeoyl)-glucoside: 655, 331, 150). For more details of the MS/MS experimental conditions see Ref. [42].

2.3. FT-IR spectroscopy

A FT-IR spectrometer Foss (Hillerød, Denmark) WineScan FT 120 specifically designed by the manufacturer for wine analysis was used to generate the FT-IR spectra. This instrument employs a Michelson interferometer to generate the spectra and has been employed since 1996 with ready-to-use must and wine calibrations provided by the manufacturer for general enological parameters (not for anthocyanins) [29]. FT-IR spectra of samples were registered in April 2006, thus young samples of 2005 vintage were just bottled, whereas young samples of vintage 2004 had passed one year within bottle. Wine samples are filtered by the instrument prior to analysis and thermostatted at 40 °C to obtain reproducible IR spectra.

Samples (~30 mL) were pumped through a CaF_2-lined cuvette (optical path length 37 μm), which is housed in the heater unit of the instrument. FT-IR spectra of wine samples were acquired in the frequencies region 5012–926 cm^{-1} (instrument manufacturer designates this region as "pin numbers" 240–1299) at 4 cm^{-1} intervals. Certain ranges of frequencies are not taken into account to prevent noise being included in the calculations. The following spectral ranges are used to select the frequencies: 965–1543, 1717–1813 and 2701–2971 cm^{-1} (Fig. 1). The two regions 1543–1717 and 2971–3742 cm^{-1} contain strong water absorption bands and the regions 1813–2701 cm^{-1} and from 3627 cm^{-1} onwards are eliminated because they contain very little useful information.

2.4. Multivariate data analysis

PCA of the 12 anthocyanin reference concentrations obtained by HPLC–DAD and PCA of the spectral data obtained by FT-IR, both analyses of the 158 wine samples within the calibration set were

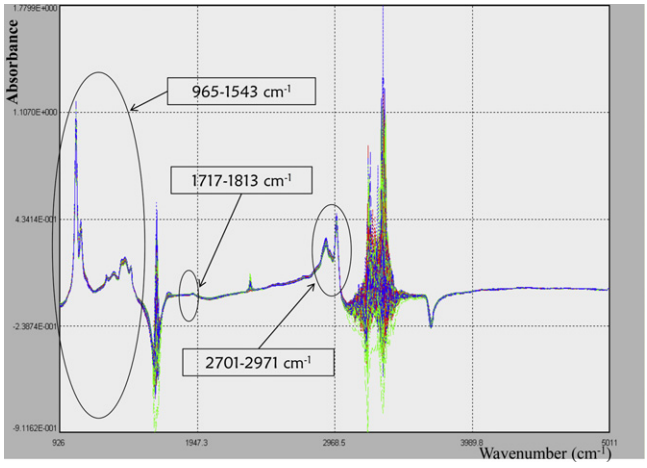

Fig. 1. Spectral ranges used to select the useful frequencies.

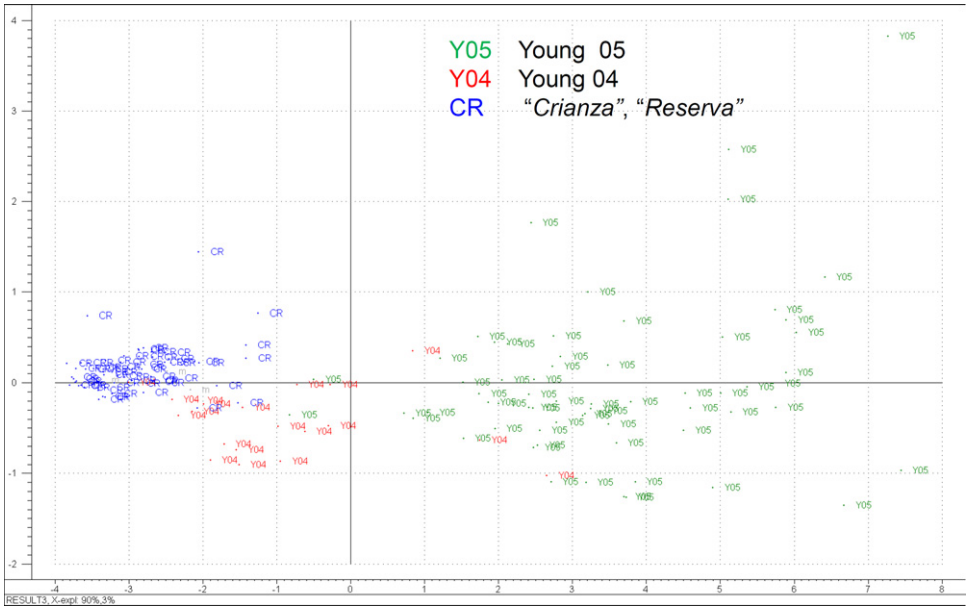

Fig. 2. Plot of PC1 vs PC2, obtained from PCA of the reference anthocyanin concentrations for the 153 samples of wine (without outliers).

carried out using the software Unscrambler (version 9.2, CAMO Process AS, Oslo, Norway).

The chemometric tool used in the WineScan calibration software FT 120 v 2.2.2. for quantitative determinations is based on Partial Least Squares Regression (PLS-R). Prior to PLS regression, the appropriate wavelength ranges (individual frequencies (cm^{-1}) or small groups of consecutive frequencies) for each anthocyanin were selected using the variable selection tool of the FT 120 software based on the correlation of the variation in each frequency in the spectra to the reference results over the samples. After that, the PLS-R calculation compresses the selected predictor variables into PLS-R factors.

In this study the optimum number of PLS-R factors was selected for each anthocyanin as that corresponding to the first local minimum of cross validation error (CVE) values. Thus, the overfitting of the model is avoided and the maintenance of a low cross validation error (CVE) in each calibration is insured.

The performance of the calibration models were evaluated by the Standard Error of Calibration (SEC), the correlation coefficient (R^2) and the errors of cross validation (CVE). SEC indicates the accuracy with which the reference value can be predicted for the calibration sample and was calculated as:

$$\mathrm{SEC} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \quad (1)$$

where N is the number of samples, y_i is the reference value for sample i, and x_i is the predicted value for sample i.

Cross validation was done keeping out successive groups of samples from the calibration set (25% of the total number of calibration samples at a time), and using these subsets for prediction on the basis of the rest of the samples, this procedure was repeated four times so all samples were included in validation sets. This procedure provides a good estimate of how accurately the calibration may be expected to work with an independent sample set. Error of cross validation was estimated as follows:

$$\mathrm{CVE} = \sqrt{\frac{\sum_{s=1}^{S}\sum_{i=1}^{n}(y_{is} - x_{is})^2}{N}} \quad (2)$$

where S is the number of Subsets, n is the number of samples in a given group, N is the total number of samples, y_{is} is the reference value for sample i and subset s, and x_{is} is the predicted value for sample i and subset s.

An additional validation using an external validation sample set was also performed, comparing the values predicted using the model built with the calibration sample set with the values obtained with the reference method.

3. Results and discussion

3.1. Reference results

Wine spectra are extremely complex because they are caused by many bands of different compounds and many of them are overlapping. Therefore, it is necessary to use advanced mathematical techniques to generate individual calibration equations for the studied parameters: individual anthocyanins and three sums of anthocyanins (non-acylated, acetylated and coumaroylated anthocyanins).

Prior to the building calibrations, Principal Component Analysis (PCA) of anthocyanin concentrations and spectral data were used to find outlier samples in order not to be included in the calibration treatment, obtaining robust calibration models. Thus, 5 samples were considered as outliers and were deleted from the later treatments, in all cases corresponding to wines from other geographical regions (not *Rioja*) or from a different grape cultivar (not *Tempranillo*). The homogeneity of the samples included into the calibration greatly conditions the prediction errors attained. After these deletions, a new PCA of the anthocyanin concentrations was built and a differentiation among types of wine was observed, as it is shown in Fig. 2. The first two Principal Components (PCs) describe the 93% of the total variability of the data. The wines appear distributed into three groups (*young* of 2005, *young* of 2004 and *crianza* and *reserva*) along PC1 axe, whereas little differentiation is shown by PC2 that only explains the 3% of the total variance.

Different concentrations (Table 2) and chromatographic profiles (Fig. 3) were obtained for *young*, *crianza* and *reserva* wines. Anthocyanins concentration values decrease as the wine ageing occurs (in fact, the first PC in Fig. 2 can be interpreted as the total amount of anthocyanins in wine which decreases as wine ageing increases). For *reserva* wines, many new derived anthocyanic pigments [43] are formed, coelluting and originating a characteristic

Table 2
Summary of mean and range of anthocyanin concentrations (mg/L equivalents of Malvidin-3-*O*-glucoside) in the different types of wine analyzed (n = 158 samples).

Anthocyanins	*Young*		*Crianza*		*Reserva* and *gran reserva*	
	Mean ± SD (mg/L)	Range	Mean ± SD (mg/L)	Range	Mean ± SD (mg/L)	Range
Delphinidin-3-*O*-glucoside	26 ± 10	4.9–53.8	7.8 ± 6.7	1.6–31.3	3.9 ± 3.7	0.5–20.3
Cyanidin-3-*O*-glucoside	2.34 ± 0.94	0.4–4.5	0.71 ± 0.72	0.1–3.6	0.38 ± 0.32	0.1–1.6
Petunidin-3-*O*-glucoside	29 ± 12	5.9–57.2	7.9 ± 5.8	1.8–25.8	4.0 ± 3.6	0.4–19.3
Peonidin-3-*O*-glucoside	8.1 ± 3.7	1.4–19.9	2.1 ± 1.7	0.3–7.6	1.1 ± 1.0	0.1–4.9
Malvidin-3-*O*-glucoside	121 ± 49	18.8–234.5	31 ± 18	6.7–76.7	16 ± 12	1.6–66.6
Petunidin-3-*O*-(6-*O*-acetyl)-glucoside	1.8 ± 1.3	0.2–9.6	0.39 ± 0.33	0.0–1.7	0.18 ± 0.22	0.0–0.9
Peonidin-3-*O*-(6-*O*-acetyl)-glucoside	0.90 ± 0.93	0.1–6.6	0.30 ± 0.15	0.1–0.8	0.24 ± 0.15	0.1–0.8
Malvidin-3-*O*-(6-*O*-acetyl)-glucoside	7.9 ± 8.8	0.7–65.7	1.76 ± 0.85	0.3–3.5	1.0 ± 1.1	0.1–6.4
Malvidin-3-*O*-(6-*O*-caffeoyl)-glucoside	1.00 ± 0.57	0.1–2.9	0.27 ± 0.21	0.0–1.1	0.18 ± 0.12	0.0–0.6
Petunidin-3-*O*-(6-*O*-p-coumaroyl)-glucoside	3.2 ± 1.5	0.5–7.0	0.83 ± 0.55	0.2–2.4	0.40 ± 0.32	0.0–1.5
Peonidin-3-*O*-(6-*O*-p-coumaroyl)-glucoside	1.97 ± 0.93	0.3–4.2	0.47 ± 0.31	0.1–1.3	0.33 ± 0.21	0.1–1.1
Malvidin-3-*O*-(6-*O*-p-coumaroyl)-glucoside	14.2 ± 6.0	2.2–28.4	3.6 ± 2.0	0.7–8.2	1.8 ± 1.3	0.1–5.9
ΣAnthocyanins non-acylated	187 ± 73	31.4–359.8	50 ± 32	10.7–145.0	25 ± 21	2.7–112.9
ΣAnthocyanins acetylated	11 ± 11	1.1–81.9	2.5 ± 1.2	0.5–5.9	1.4 ± 1.4	0.2–7.9
ΣAnthocyanins coumaroylated	19.4 ± 8.3	3.0–38.2	4.9 ± 2.9	1.0–12.0	2.6 ± 1.7	0.7–8.5

hump in the chromatogram. The most abundant anthocyanin in all the samples was Malvidin-3-glucoside, followed by Petunidin-3-glucoside and Delphinidin-3-glucoside in order of importance. Moreover, the concentrations of coumaroylated anthocyanins are bigger than acetylated ones.

The 153 samples employed in the calibration encompass the characteristics of a broad range of red wine of *Rioja*, thus enough variation in anthocyanin levels (Table 2) has been introduced to allow a suitable calibration.

3.2. FT-IR results

After selecting the optimum number of frequencies as predictor variables for each response variable (12 individual anthocyanin and the 3 sums [non-acylated, acetylated and coumaroylated anthocyanins]), one individual PLS calibration model was built for each response variable to be predicted using the 153 sample set. Table 3 collects the values of the Standard Error of Calibration (SEC), the correlation coefficient (R^2) and the errors of cross validation (CVE)

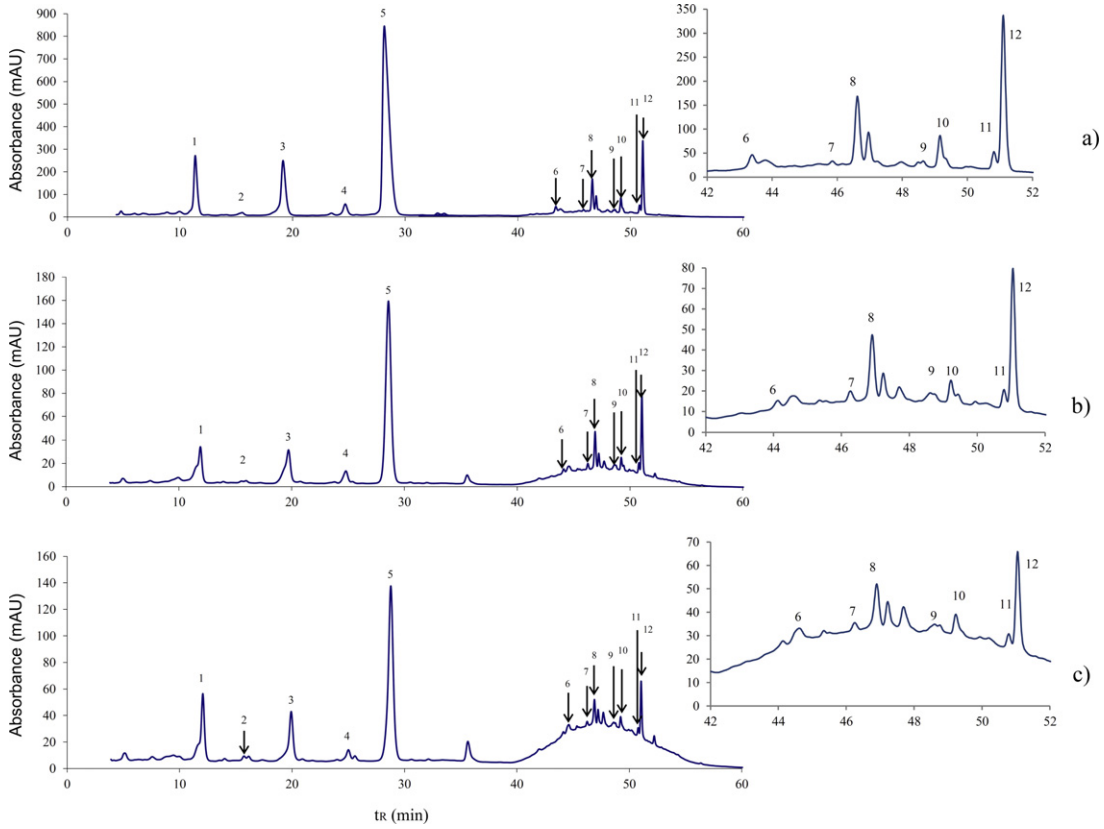

Fig. 3. Typical HPLC chromatograms at 530 nm for *Rioja* red wines: (a) *young*, (b) *crianza* and (c) *reserva*. 1: Delphindin-3-*O*-glucoside; 2: Cyanidin-3-*O*-glucoside; 3: Petunidin-3-*O*-glucoside; 4: Peonidin-3-*O*-glucoside; 5: Malvidin-3-*O*-glucoside: 6: Petunidin-3-(6-*O*-acetyl)-glucoside; 7: Peonidin-3-(6-*O*-acetyl)-glucoside; 8: Malvidin-3-(6-*O*-acetyl)-glucoside; 9: Malvidin-3-(6-*O*-caffeoyl)-glucoside; 10: Petunidin-3-(6-*O*-*p*-coumaroyl)-glucoside; 11: Peonidin-3-(6-*O*-*p*-coumaroyl)-glucoside; 12: Malvidin-3-(6-*O*-*p*-coumaroyl)-glucoside.

Table 3
Mean concentrations and range (mg/L equivalents of Malvidin-3-*O*-glucoside) of individual anthocyanins and sums of anthocyanins of the 153 samples included in the calibration models. Optimum number of PLS-R factors. SEC (mg/L) and R^2 of the calibration models and CVE (mg/L, %) obtained in cross validation.

Anthocyanins	Reference data		Calibration			Cross validation	
	Mean (mg/L)	Range	Optimum number of PLS factors	SEC (mg/L)	R^2	CVE (mg/L)	%CVE
Delphinidin-3-glucoside	16.8	0.5–53.8	10	4.2	0.90	5.0	29
Cyanidin-3-glucoside	1.5	0.1–4.5	10	0.4	0.87	0.5	32
Petunidin-3-glucoside	18.5	0.4–57.2	13	4.4	0.91	5.0	27
Peonidin-3-glucoside	5.1	0.1–19.9	12	1.7	0.85	2.1	41
Malvidin-3-glucoside	76.2	1.6–234.5	9	17.9	0.92	18.7	26
Petunidin-3-(6-*O*-acetyl)-glucoside	1.0	0.1–3.3	10	0.3	0.87	0.4	37
Peonidin-3-(6-*O*-acetyl)-glucoside	0.5	0.1–2.5	15	0.2	0.64	0.2	48
Malvidin-3-(6-*O*-acetyl)-glucoside	4.0	0.1–20.5	10	1.5	0.81	1.6	43
Malvidin-3-(6-*O*-caffeoyl)-glucoside	0.6	0.1–2.9	9	0.3	0.80	0.3	44
Petunidin-3-(6-*O*-p-coumaroyl)-glucoside	2.0	0.1–7.0	12	0.6	0.90	0.6	32
Peonidin-3-(6-*O*-p-coumaroyl)-glucoside	1.2	0.1–4.2	12	0.4	0.87	0.4	35
Malvidin-3-(6-*O*-p-coumaroyl)-glucoside	8.8	0.1–28.4	10	2.1	0.92	2.4	28
ΣAnthocyanins non-acylated	118.1	2.7–359.8	12	25.7	0.93	29.5	25
ΣAnthocyanins acetylated	5.5	0.2–26.3	13	1.9	0.84	2.0	38
ΣAnthocyanins coumaroylated	12.1	0.7–38.2	12	2.9	0.92	3.1	26

obtained and the optimum number of factors employed in each calibration. The low differences attained between SEC and CVE values show that the built models are near to the real situation of the sample set. Therefore, it seems that the model is not overfitted.

In general, as expected, anthocyanins with a higher concentration level are better predicted, showing lower %CVE values (25–30%) and the analytical calibration of compounds present in low concentrations is more affected by other matrix compounds, present in higher concentrations and with identical or very close IR absorption bands. This is the case of Peonidin-3-*O*-glucoside, Peonidin-3-(6-*O*-acetyl)-glucoside, Malvidin-3-(6-*O*-acetyl)-glucoside and Malvidin-3-(6-*O*-caffeoyl)-glucoside. However, other anthocyanins, such as Cyanidin-3-*O*-glucoside and Petunidin-3-(6-*O*-*p*-coumaroyl)-glucoside, show better CVE values in spite of their low concentrations.

CVE values obtained are slightly better, in general, than the results of Soriano et al. [22]. For example, Soriano et al. reach SEC of 19.72, 4.81, 4.27, 6.89 and 4.29 mg/L against 17.9, 4.4, 4.2, 1.9 and 2.9 mg/L, for Malvidin-3-glucoside, Petunidin-3-glucoside, Delphinidin-3-glucoside, total acetylated and total coumaroylated anthocyanins, respectively; the existent differences showing that as more homogeneous group of samples are selected for calibration purposes, CVE values are lower. Whereas the study of Soriano et al., included *young* red wines from six different geographical regions and seven cultivars, this paper focuses on mainly a variety (*Tempranillo*) and a unique Protected Designation of Origin (*Rioja*). Our calibration set has the heterogeneity introduced by the different stages of ageing of wines (*young*, *crianza* and *reserva*). If a more homogeneous sample set were used (only *young* wines) to generate calibrations these errors would be probably reduced.

3.3. External validation

An estimation of prediction errors closer to the real situation when analyzing unknown wine samples was performed by an external validation using some wines provided by the wineries and other wines purchased in the market (Table 1). All these samples were analyzed by the FT-IR equipments of three wineries after the implementation of the calibrations built. Fig. 4 shows the reference values by HPLC–DAD and the FT-IR predicted concentration values (mg/L equivalents of Mv-3-*O*-glc) for five anthocyanins and for the sums of non-acylated, acetylated and coumaroylated anthocyanins obtained in one of the three wineries.

Table 4
Reference concentration values (mg/L) (mean ± standard deviation) and standard prediction errors (%SEP) of samples of external validation.

Anthocyanins	External validation					
	Young 2005		*Crianza*		*Reserva* and *gran reserva*	
	Reference value (mg/L) (mean ± SD)	%SEP	Reference value (mg/L) (mean ± SD)	%SEP	Reference value (mg/L) (mean ± SD)	%SEP
Delphinidin-3-glucoside	26.4 ± 9.3	22	5.6 ± 2.6	65	1.72 ± 0.43	259
Cyanidin-3-glucoside	2.09 ± 0.56	23	0.55 ± 0.18	51	0.32 ± 0.12	150
Petunidin-3-glucoside	30 ± 12	21	6.0 ± 2.4	49	1.68 ± 0.41	247
Peonidin-3-glucoside	8.5 ± 2.7	18	1.81 ± 0.59	80	0.40 ± 0.15	456
Malvidin-3-glucoside	121 ± 48	22	26.9 ± 6.4	39	6.5 ± 1.8	316
Petunidin-3-(6-*O*-acetyl)-glucoside	1.54 ± 0.76	32	0.32 ± 0.09	92	0.10 ± 0.05	278
Peonidin-3-(6-*O*-acetyl)-glucoside	0.59 ± 0.23	63	0.24 ± 0.08	81	0.18 ± 0.04	127
Malvidin-3-(6-*O*-acetyl)-glucoside	5.8 ± 2.4	28	1.59 ± 0.37	66	0.31 ± 0.11	527
Malvidin-3-(6-*O*-caffeoyl)-glucoside	1.08 ± 0.44	34	0.21 ± 0.06	100	0.05 ± 0.02	540
Petunidin-3-(6-*O*-p-coumaroyl)-glucoside	3.7 ± 1.5	27	0.86 ± 0.28	58	0.16 ± 0.03	268
Peonidin-3-(6-*O*-p-coumaroyl)-glucoside	1.90 ± 0.70	21	0.36 ± 0.11	60	0.10 ± 0.04	371
Malvidin-3-(6-*O*-p-coumaroyl)-glucoside	15.6 ± 5.0	17	3.50 ± 0.89	54	1.10 ± 0.19	206
ΣAnthocyanins non-acylated	187 ± 71	17	41 ± 12	35	10.7 ± 2.7	265
ΣAnthocyanins acetylated	7.9 ± 3.4	18	2.15 ± 0.49	59	0.58 ± 0.06	268
ΣAnthocyanins coumaroylated	21.3 ± 7.1	15	4.7 ± 1.3	46	1.36 ± 0.25	315

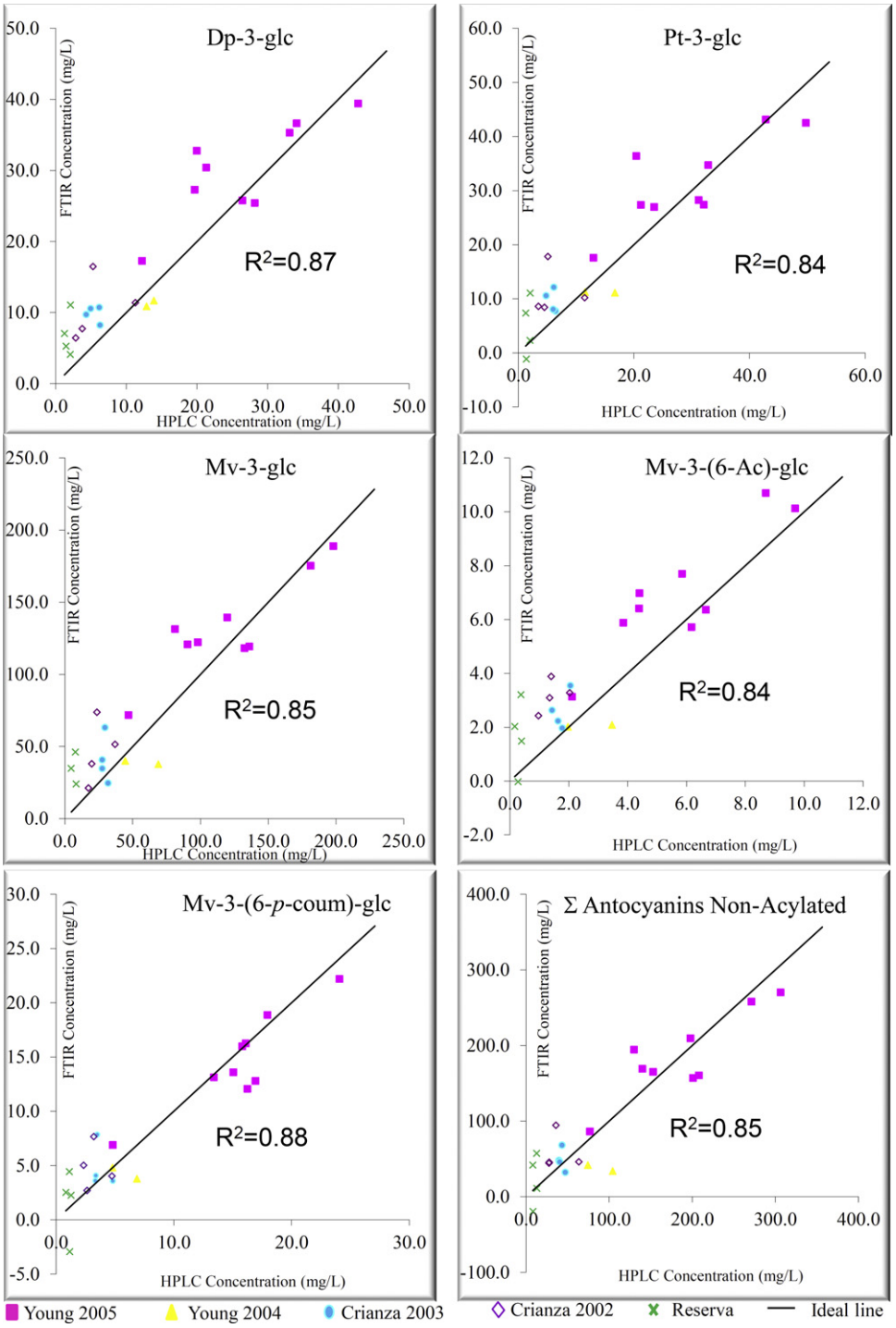

Fig. 4. Comparison between reference concentration values by HPLC–DAD and FT-IR predicted concentration values (mg/L equivalents of Mv-3-*O*-glc) for five anthocyanins and for the sums of non-acylated, acetylated and coumaroylated anthocyanins. R^2 values of the regression lines taking into account all the samples of the external validation set are shown, together with the ideal line (predicted value = reference value).

As it can be seen in the figures, there is a correlation between reference and IR predicted data in the validation external set. Therefore, it seems that the models work properly in the studied range of concentrations.

The best results were obtained for *young* wine samples recently bottled, as it was expected due to their higher anthocyanin concentrations. The relative Standard Error of Predictions (SEP, see Table 4) in the cases of *reserva* and *gran reserva* wines (>100%), *crianza* wines (40–70%) were too high to recommend this method for the analysis of these kinds of wines. Thus, only for *young* wines recently bottled, the predictions errors were enough low (15–30%) for major anthocyanins and for the sums of anthocyanins to make recommendable the FT-IR prediction for routine analysis at wineries. Moreover, these relatively high errors compared with other instrumental techniques such as HPLC are acceptable on the basis of the fast and direct analysis of wine that this technique allows;

it is strongly advisable that the winery users were aware of this fact and of the need of a robust external validated calibration of the technique.

4. Conclusions

FT-IR with PLS-R offers some advantages when compared against conventional methods of analysis as a high-through-put tool for routinely screening of wines at wineries. This study shows that the built calibration model is only applicable to *young* wines recently bottled and not to *young* wines after long periods being bottled or aged wines. A more specific calibration for *young* wines recently bottled would improve these results. The similarity between the obtained results by the model in the internal and external validation leads to assert that the model is robust and effective. However, winery users must be aware of the need of a robust external validated calibration of the technique in order to maintain the prediction errors within controlled limits.

Acknowledgements

This research was supported by the Agricultural Department of Basque Government. Miriam Romera Fernández and Sergio Garmón Lobato thank Universidad del País Vasco/Euskal Herriko Unibertsitatea and Gobierno Vasco/Eusko Jaurlaritza, respectively, for their Ph.D. grants. Authors also thank to the wineries Faustino, Cune, Coto de Rioja, Torre de Oña and Unión de Cosecheros de Labastida, the association of wineries ABRA and Foss, for their collaboration.

References

[1] R.R. Gangidi, A. Proctor, F.W. Pohlman, J.F. Meullenet, J. Food Sci. 70 (2005) c397–c400.
[2] M. Pla, P. Hernández, B. Ariño, J.A. Ramirez, I. Diaz, Food Chem. 100 (2007) 165–170.
[3] M. Uddin, E. Okazaki, H. Fukushima, S. Turza, Y. Yumiko, Y. Fukuda, Food Chem. 69 (2006) 491–495.
[4] H. Yang, J. Irudayaraj, J. Am. Oil Chem. Soc. 77 (2000) 291–295.
[5] L.P. Meagher, S.E. Holroyd, D. Illingworth, F. van de Ven, S. Lane, J. Agric. Food Chem. 55 (2007) 2791–2796.
[6] D. Bertrand, P. Robert, W. Loisel, J. Sci. Food Agric. 36 (1985) 1120–1124.
[7] S.R. Delwiche, R.A. Graybosch, J. Cereal Sci. 35 (2002) 29–38.
[8] Z. Seregely, T. Deak, G.D. Bisztray, Chemometr. Intell. Lab. Syst. 72 (2004) 195–203.
[9] A.M.K. Pedro, M.M.C. Ferreira, Anal. Chem. 77 (2005) 2505–2511.
[10] J.D. Kelly, G. Downey, J. Agric. Food Chem. 53 (2005) 3281–3286.
[11] C.J. Bevin, A.J. Fergusson, W.B. Perry, L.J. Janik, D.J. Cozzolino, Agric. Food Chem. 54 (2006) 9713–9718.
[12] S. Engelhard, H.G. Lohmannroben, F. Schael, Appl. Spectrosc. 58 (2004) 1205–1209.
[13] A. Giunchi, A. Berardinelli, L. Ragni, A. Fabbri, F.A. Silaghi, J. Food Eng. 89 (2008) 142–148.
[14] G.F. Baumgarten, S. Afr, J. Enol. Vitic. 8 (1987) 75–77.
[15] R. Schindler, R. Vonach, B. Lendl, R. Kellner, Fresen. J. Anal. Chem. 362 (1998) 130–136.
[16] C.D. Patz, A. David, K. Thente, P. Kürbel, H. Dietrich, Vitic. Enol. Sci. 54 (1999) 80–87.
[17] B.M. Jespersen, L. Munck, in: D.W. Sun (Ed.), Infrared Spectroscopy for Food Quality Analysis and Control, Elsevier, Dublin, Ireland, 2009, pp. 377–386.
[18] W.F. McClure, J. Review, Near Infrared Spectrosc. 11 (2003) 487–518.
[19] M. Dubernet, M. Dubernet, Revue Française dïOenologie 181 (2000) 10–13.
[20] S.A. Kupina, A.J. Shrikhande, Am. J. Enol. Vitic. 54 (2003) 131–134.
[21] R. Lletí, E. Meléndez, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Anal. Chim. Acta 544 (2005) 60–70.
[22] A. Soriano, P.M. Pérez-Juan, A. Vicario, J.M. González, M.S. Pérez-Coello, Food Chem. 104 (2007) 1295–1303.
[23] D. Cozzolino, M.J. Kwiatkowski, R.G. Dambergs, W.U. Cynkar, L.J. Janik, G. Skouroumounis, M. Gishen, Talanta 74 (2008) 711–716.
[24] M. Dubernet, M. Dubernet, V. Dubernet, S. Coulomb, M. Lerch, I. Traineau, Bull. OIV 15 (2000) 839–840.
[25] S.K. Anderson, P.W. Hansen, H.V. Anderson, in: J.M. Chalmers, P.R. Griffiths (Eds.), Handbook of Vibrational Spectroscopy, John Wiley & Sons Ltd., Chichester, United Kingdom, 2002, pp. 2682–3672.
[26] S. Macho, Thesis, Universidad Rovira i Virgili, Barcelona, 2002.
[27] C.D. Patz, A. Blieke, R. Ristow, H. Dietrich, Anal. Chim. Acta 513 (2004) 81–89.
[28] A. Urtubia, J.R. Pérez-Correa, M. Meurens, E. Agosin, Talanta 64 (2004) 778–784.
[29] M. Urbano-Cuadrado, M.D. Luque de Castro, P.M. Pérez-Juan, J. García-Olmo, M.A. Gómez-Nieto, Anal. Chim. Acta 527 (2004) 81–88.
[30] M. Urbano-Cuadrado, M.D. Luque de Castro, J. García-Olmo, M.A. Gómez-Nieto, Talanta 66 (2005) 218–224.
[31] R.A. Cocciardi, A.A. Ismail, J. Seedman, J. Agric. Food Chem. 53 (2005) 2803–2809.
[32] L. Munck, L. Norgaard, S.B. Engelsen, R. Bro, C.A. Andersson, Chemometr. Intell. Lab. Syst. 44 (1998) 31–60.
[33] I.L. Francis, P. Høj, R. Dambergs, M. Gishen, M.B. Lopes, P. Godden, P. Henscke, E. Waters, M. Herderich, I. Pretorius, Wine Ind. J. 20 (3) (2005) 12–18.
[34] M.D. Rivero-Pérez, P. Muñiz, M.L. González-Sanjosé, Food Chem. Toxicol. 46 (2008) 2815–2822.
[35] F. Alén-Ruiz, M.S. García-Falcón, M.C. Pérez-Lamela, E. Martínez-Carballo, J. Simal-Gándara, Food Chem. 113 (2009) 53–60.
[36] A. Versari, R.B. Boulton, G.P. Parpinello, Ital. J. Food Sci. 18 (2006) 423–432.
[37] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA and PLS), Umetrics AB, Umeå, Sweden, 1999.
[38] T. Næs, T. Isaksson, T. Fearn, T. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR Publications, Chichester, United Kingdom, 2004.
[39] K.H. Esbensenin, Multivariate Data Analysis-In Practice. An Introduction to Multivariate Data Analysis and Experimental Design, Camo AS, Oslo, Norway, 2006.
[40] J.L. Moreira, L. Santos, Anal. Chim. Acta 513 (2004) 263–268.
[41] B. Abad-García, L.A. Berrueta, S. Garmón-Lobato, B. Gallo, F. Vicente, J. Chromatogr. A 1216 (2009) 5398–5415.
[42] B. Abad-García, L.A. Berrueta, S. Garmón-Lobato, B. Gallo, F. Vicente, J. Mass Spectrom. 44 (2009 b) 1017–1025.
[43] M. Monagas, B. Bartolomé, C. Gómez-Cordovés, Crit. Rev. Food Sci. 45 (2005) 85–118.