Application of FT-MIR Spectroscopy for Fast Control of Red Grape Phenolic Ripening

Sandra Fragoso, Laura Aceña, Josep Guasch, Olga Busto, and Montserrat Mestres*

Group of Analytical Chemistry of Wine and Food, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Faculty of Oenology of Tarragona, Rovira i Virgili University, 43007 Tarragona, Spain

ABSTRACT: The content of phenolic compounds determines the state of phenolic ripening of red grapes and is a key criterion in setting the harvest date to produce quality red wines. In this study, the feasibility of Fourier transform mid-infrared (FT-MIR) spectroscopy combined with partial least-squares (PLS) regression to quantify phenolic compounds is reported. The reference methods used for quantifying these compounds (which were evaluated as total phenolic compounds, total anthocyanins, and condensed tannins) were the usual ones used in cellars that employed UV–vis spectroscopy. To take into account the high natural variability of grapes when building the calibration models, fresh grapes from six varieties, at different phenolic ripening states were harvested during three vintages. Destemmed and crushed grapes were subjected to an accelerated extraction process and used as calibration standards. A total of 192 extracts (objects) were obtained, and these were divided into a training set (106 objects) and a test set (86 objects) to evaluate the predictive ability of the models. Among the different MIR regions of the extract raw spectra, those that provided the highest variability on the absorption were selected. The results showed that the best PLS regression model was the one obtained when working in the region of $1168-1457$ cm^{-1} because it gave the most accurate and robust prediction for total phenolic compounds (RMSEP % = 4.3 and RPD = 4.5), total anthocyanins (RMSEP % = 5.9 and RPD = 3.5), and condensed tannins (RMSEP % = 5.8 and RPD = 3.8). Therefore, it can be concluded that FT-MIR spectroscopy can be a fast and reliable technique for monitoring the phenolic ripening in red grapes during the harvest period.

KEYWORDS: FT-MIR spectroscopy, phenolic compounds, PLS regression, phenolic ripening

■ INTRODUCTION

The study of phenolic compounds has taken on a special significance in recent years due to their contribution to the organoleptic properties of red wine. The two most important classes of phenolic compounds found in grapes are anthocyanins and tannins.[1] Anthocyanins, located in grape skins, are responsible for the color of red grapes.[2] Tannins, located in grape seeds and skins, are mainly responsible for mouthfeel properties such as astringency and body of wine.[3] Moreover, tannins are important because they affect the development of color during wine aging. The concentration of the phenolic compounds increases during grape ripening, although it is affected by many factors including grape variety,[4] climatic and soil conditions ("terroir"),[5,6] and viticulture techniques and enological treatments.[7,8] Therefore, the determination of this concentration, which has a direct relationship with the phenolic ripening state, is a key criterion in setting the harvest date to produce quality red wines.

The usual analytical methods used to quantify the phenolic parameters in grapes are based on spectrophotometric measurements at 520 nm for total anthocyanins[9] and at 280 nm for total phenolic compounds[2] and condensed tannins.[10] These methods are generally used in cellars because they are sufficiently simple and precise. However, during grape ripening controls, a large number of samples have to be analyzed daily and, then, these methods become tedious and time-consuming.

The use of spectrometric techniques, such as Fourier transform mid-infrared (FT-MIR) spectroscopy, has recently emerged as a powerful analytical tool that allows the fast and simultaneous analysis of several parameters in a large number of samples. Due to the great amount of information about the sample composition that each FT-MIR spectrum provides, it is necessary to use of chemometrics tools to make the most of its potential both for qualitative analysis and for quantitative analysis. Partial least-squares (PLS) regression is a multivariate calibration method that is particularly useful when we need to predict a set of dependent variables from a large set of independent variables (i.e., predictors).[11]

FT-MIR equipments have already been demonstrated to be suitable for routine qualitative analysis and process control in wineries by analyzing raw materials,[12] monitoring the fermentation[13,14] or determining some of the main compounds in wine such as ethanol, organic acids, or sugars.[15−17] In regard to the analysis of phenolic compounds, there are also some applications of FT-MIR spectroscopy related to the quantification of these compounds in white[18] and red wines.[19−24] However, studies to determine the concentration of phenolic compounds in red grape homogenates have been developed using near-infrared spectroscopy (NIR).[25,26] Moreover, in this study frozen samples were used for developing the calibration models, and this sample conservation treatment affects the accuracy of the predictive results as we demonstrated in previous studies,[27] mainly due to the effect of the rupture of the cell membrane of the skins by the ice crystals, which could affect the precision of the extraction step.

Therefore, the aim of this study was to evaluate the potential of FT-MIR, combined with PLS multivariate calibration, for quantifying phenolic compounds in red grape. With this objective, the phenolic contents (evaluated as total phenolic compounds, total

anthocyanins, and condensed tannins) of different fresh red grapes at different degrees of maturity were spectrophotometrically determined according to the methods usually employed in cellars, and the results obtained were correlated with FT-MIR response. Then, the models obtained were used to predict the concentration of new samples. From these predictions we evaluated the reliability of this technique for providing the value of some phenolic parameters commonly used in cellars, which would allow this technique to be used to monitor phenolic ripening in red grapes.

■ MATERIALS AND METHODS

Sample Preparation. Fresh grape berries of six red varieties (Merlot, Tempranillo, Syrah, Garnacha, Cariñena, and Cabernet sauvignon), at different states of phenolic ripening (from veraison until harvest) and from three different vintages (2007, 2008, and 2009) were obtained from the experimental vineyard of the Faculty of Oenology of Tarragona (Rovira i Virgili University, Spain). To ensure a representative sampling, 200 berries from central rows and alternative vines, combining sunlight and shadow and different parts of the cluster (upper, middle, and bottom) were chosen.

Grapes were destemmed and ground at room temperature using an Ultra-Turrax high-speed homogenizer at 24000 rpm for 2 min to get a smooth paste. The phenolic compounds of each sample were extracted by using the method previously optimized in our laboratory[27] that provides a recovery of total phenolic compounds of >95%. Therefore, 50 g of paste was macerated during 15 min at 40 °C in a hydroalcoholic acid solution (85 mL of HCl 1% v/v + 15 mL of ethanol 96%) under constant agitation. Then the sample was centrifuged for 10 min at 8000 rpm, and the precipitate was resuspended in 50 mL of extracting solution and recentrifuged. Finally, the supernatants of both centrifugations were joined, and the volume was completed to 200 mL with the extracting solution.

Instrumentation. The grapes were ground using a high-speed homogenizer Ultra-Turrax T-18 (IKA, USA) equipped with a S18N-19G rotating shaft. Sample centrifugation was carried out by a Hettich Universal 32 R centrifuge (Tuttlingen, Germany). The reference analytical measurements were performed by using a Thermo Spectronic ultraviolet−visible spectrophotometer model Helios γ (Thermo Electron Corp., Cambridge, U.K.). All spectra of the extracts were collected using a FT-MIR Nexus (Thermo, USA), equipped with a deuterated triglycine sulfate detector (DTGS). The instrument was connected to a TDI Bacchus (Gavà, Spain) autosampler. The software package OMNIC version 6.2 from Thermo Nicolet was used for spectra acquisition. The software used for data analysis and calibration was the Unscrambler package (version 9.0, CAMO ASA, Norway).

Reagents and Standards. The standards of malvidin-3-glucoside (purity ≥ 90%) and (+)-catechin (purity ≥ 96%) were supplied by Fluka (Madrid, Spain). Gallic acid monohydrate (99.5%), tannic acid (99.9%), and ammonium sulfate (99.5%) were supplied by Scharlab (Barcelona, Spain). Methyl cellulose (M-0387) was supplied by Sigma Aldrich (Madrid, Spain). The rest of the chemicals used for the study were of analytical reagent grade and supplied by Scharlab.

Reference Analytical Measurements. The phenolic ripening of grapes was evaluated by three parameters: total phenolic compounds, total anthocyanins, and condensed tannins. The reference methods used for quantifying them were based on UV−vis spectroscopy.

Total Phenolic Compounds Content (TPC). This value was determined by measuring at 280 nm the absorbance of extract sample, in a 10 mm quartz cuvette, previously diluted 50 times in deionized water[2] and using a calibration line built with gallic acid monohydrate as standard at six different concentrations in the range of 2.2−18.0 mg L^{-1}. The total

phenolic compounds content was expressed in milligrams of gallic acid per kilogram of grape.

Total Anthocyanins Content (TA). This parameter was determined by measuring the absorbance at 520 nm of the extract, in a 10 mm plastic cuvette, previously diluted 25 times with 0.1 M HCl[9] to get a pH close to 1.0 and using a calibration line built using malvidin-3-glucoside as standard at six different concentrations in the range of 2.4−20.0 mg L^{-1}. The total anthocyanin content was expressed in milligrams of malvidin-3-glucoside per kilogram of grape.

Condensed Tannins Content (CT). The quantification of these phenolic compounds was carried out by using the indirect method of precipitation with methyl cellulose[10] and using a calibration line built with (+)-catechin as standard at six different concentrations in the range of 19.2−76.9 mg L^{-1}. This content was expressed in milligrams of (+)-catechin per kilogram of grape.

Mid-Infrared Scanning Spectra. The MIR spectra of the samples were collected using a FT-MIR Nexus (Thermo, USA), connected to a TDI Bacchus autosampler, which was equipped with an online and automatic system of sample filtration (stainless steel filter of food grade with a pore size of 50 μm, Teflon coated). The FT-MIR spectrum acquisition takes only 30 s per sample.

All spectra were averaged from 32 scans and collected in absorbance mode, at 4 cm^{-1} spectral resolution, on the 979−2989 cm^{-1} wavenumber range (MIR vibrational zone). To eliminate the possible equipment drift over time, every 10 h, the equipment collected an environmental spectrum (considering water vapor together with CO_2) and redefined the background.

On the other hand, because water dominates the spectrum of aqueous samples, prior to the analysis of the grape extracts, a blank of distilled water was acquired automatically by the spectrometer to isolate the water absorption bands as done in other studies.[28]

Enrichment Experiments. During grape ripening, changes in the concentration of chemical compounds different from the phenolic compounds, such as sugars or organic acids, take place. Therefore, it is difficult to determine which compound or compounds are responsible for the changes observed in the spectral response. To locate the spectral regions associated with phenolic compounds we compared a grape extract spectrum with the spectra from the same grape extract enriched with four phenolic compounds of different chemical structure. Thus, different concentration levels (ranging from 0.5 to 5 g L^{-1}) of gallic acid, (+)-catechin, and tannic acid, as well as 1 mg of malvidin-3-glucoside (commercial standard format), were added to different grape extracts. Because these compounds present different chemical compositions and polymerization grades, they were very helpful and appropriate for identifying the absorbance regions of FT-MIR spectra related to phenolic compounds present in grapes and, therefore, for selecting the suitable wavenumbers used on the calibration of phenolic compounds. Chemical structures of the four different phenolic compounds are shown in Figure 1.

Development and Validation of the FT-MIR Models. Chemometrics was used to perform both descriptive and quantitative analysis of the data. The descriptive analysis was done using the principal components analysis (PCA) method. PCA is used to describe the maximal quantity of information present in the data set using a small number of latent variables (i.e., not directly measured), and it is very useful to reveal possible grouping samples and to visualize the presence of outliers.[29]

Quantitative analysis was done using PLS regression. PLS regression is a method used for relating a matrix of predictor variables, **X** (i.e., spectra), and a matrix of chemical properties, **Y** (i.e., concentration values), by a linear multivariate model.[11] For this, **X** is divided into a training set to build the model and a set to evaluate the prediction ability of the model. A critical step in the model building is the selection of the number of optimal factors (latent variables) to ensure the prediction
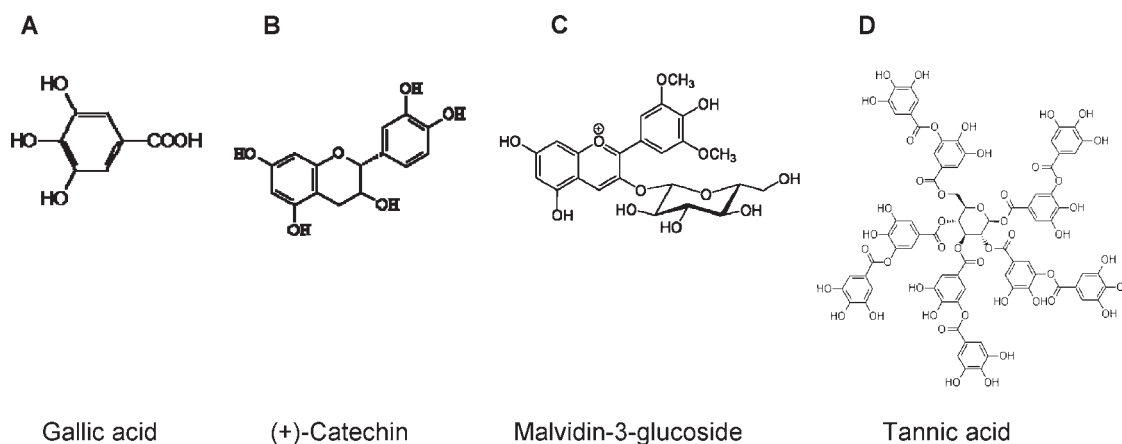
2176

dx.doi.org/10.1021/jf104039g |*J. Agric. Food Chem.* 2011, 59, 2175–2183

Figure 1. Chemical structures of the four different standards used to determine the FT-MIR spectra absorbance regions of the phenolic compounds: gallic acid (A), (+)-catechin (B), malvidin-3-glucoside (C), and tannic acid (D).

ability and to avoid overfitting. In this paper the number of optimal factors was obtained by means of leave-one-out cross-validation.[29] In that way, one object of the training set is removed and a PLS model is built with the remaining ones and, then, this model is used to predict the removed object. This procedure is repeated until all of the objects in the training set are selected for a number of given factors. The optimal number of factors in the PLS model was determined by the lowest root mean square error of cross-validation (RMSECV) (eq 1)

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_{\backslash i})^2}{N_{\text{CV}}}} \quad (1)$$

where N_{CV} is the number of cross-validation samples (i.e., $N_{\text{CV}} = N$, N being the number of samples in the training set), y_i is the reference measurement, and $\hat{y}_{\backslash i}$ is the estimated result when the model is built without sample i.

Related to the calibration step, we evaluated the model fit to the data with the root mean square error of calibration (RMSEC) expressed as a percentage (RMSEC %) (eq 2) that can be defined as the mean error of model.

$$\text{RMSEC \%} = \frac{100}{\bar{y}} \left(\sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \right) \quad (2)$$

N is the number of samples of training set, y_i is the reference value for sample i, and $\hat{y}_i$ is the predicted value for sample i.

To test the predictive accuracy of the calibration models built, the minimal root mean square error of prediction (RMSEP) was determined in percent (eq 3)

$$\text{RMSEP \%} = \frac{100}{\bar{y}} \left(\sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \right) \quad (3)$$

where N is the number of samples of the test set, y_i is the reference value for sample i, and $\hat{y}_i$ is the predicted value for sample i. The results of future predictions can then be expressed as "predicted value $\pm 2 \times$ RMSEP".[30]

Additionally, to standardize the predictive accuracy, for each model the residual predictive deviation (RPD) was calculated as the ratio

Table 1. Descriptive Statistics of Training and Test Sets[a]

phenolic parameter	training set (samples = 106)				test set (samples = 86)			
	min	max	mean	SD	min	max	mean	SD
TPC	1005	2140	1626	363	1058	2035	1623	315
TA	348	1316	810	258	392	1225	791	184
CT	984	3351	2298	595	1320	2912	2258	487

[a] TPC, total phenolic compounds in mg of gallic acid kg^{-1}; TA, total anthocyanins in mg of malvidin-3-glucoside kg^{-1}; CT, condensed tannins in mg of (+)-catechin kg^{-1}; min, minimum value; max, maximum value; SD, standard deviation.

between the standard deviation (SD) of the validation samples and the RMSEP.[31] If the RMSEP is large compared to the spread of that compound in all samples (SD), a relatively small RPD is obtained. This means that the model is not able to model suitably the variability. An RPD value ranging between 2.5 and 3.0 is considered to be poor, and the models could be applied only for very rough screening. However, generally, an RPD of >3 could be considered to be very satisfactory for prediction purposes.[12,32]

Estimation of the True Prediction Error. Because the reference values used to construct the models are not known with negligible uncertainty but instead are obtained with a standard error, the validation of multivariate calibration models using these values leads to a so-called apparent prediction error, which systematically overestimates the true prediction error. To solve this problem we used the simple correction procedure proposed by Faber et al.[33] that yields a more realistic estimate of the true prediction error (eq 4)

$$\text{MSEP}_{\text{cor}} = \text{MSEP}_{\text{app}} - \hat{\sigma}^2 \quad (4)$$

where MSEP_{cor} is the bias-corrected MSEP, MSEP_{app} is the apparent MSEP (i.e., the value that we obtained with the test set), and $\hat{\sigma}^2$ is the variance of the measurement error in the reference values.

■ RESULTS AND DISCUSSION

Reference Values. All of the samples harvested at different phenolic ripening states to consider the high natural variability of phenolic amount in grapes were extracted by the methodology described. In this way, 192 extracts were obtained. These were distributed into a training or calibration set ($n = 106$ samples) and a test or validation set ($n = 86$ samples). The reference values of the contents of total phenolic compounds (TPC), total

anthocyanins (TA), and condensed tannins (CT) were determined by using the reference analytical measurements. The reliability of these methods was previously tested (eight analyses of the same grape extract sample), and we found that the relative

Table 2. Concentration of the Phenolic Compounds in Each Variety Studied (Samples from 2007, 2008, and 2009 Vintages)[a]

grape variety	no. of samples	phenolic parameter	min	max	mean	SD
Merlot	38	TPC	1679	2140	1943	117
		TA	843	1296	1021	139
		CT	1771	2921	2435	270
Tempranillo	34	TPC	1625	1996	1812	92
		TA	651	1002	782	103
		CT	2291	3351	2879	247
Syrah	32	TPC	1648	2120	1848	121
		TA	826	1316	1024	142
		CT	1950	2828	2415	178
Cariñena	30	TPC	1005	1196	1092	47
		TA	485	759	639	70
		CT	984	1495	1358	95
Garnacha	24	TPC	1023	1295	1173	82
		TA	348	482	411	41
		CT	1448	2217	1825	260
Cabernet sauvignon	34	TPC	1328	1856	1659	150
		TA	686	883	784	58
		CT	937	2911	2453	479

[a] TPC, total phenolic compounds in mg of gallic acid kg^{-1}; TA, total anthocyanins in mg of malvidin-3-glucoside kg^{-1}; CT, condensed tannins in mg of (+)-catechin kg^{-1}; min, minimum value; max, maximum value; SD, standard deviation.

standard deviation (%rsd) was 1.21 for TCP, 0.73 for TA, and 1.27 for CT.

Table 1 summarizes the contents of the parameters studied considering all of the samples harvested and analyzed weekly during the August–October period throughout three different vintages (2007, 2008, and 2009). As can be seen, both sets had similar contents of the compounds studied, and the highest standard deviation value was associated with the CT, which could be associated with the different size (polymerization degree) of condensed tannins in grapes from its "veraison" to full ripeness.

Table 2 summarizes the concentration of the phenolic compounds evaluated considering each variety separately. We can see, for example, that Cariñena, which is characterized by its low tannic content, presented the lowest CT value or that Merlot and Syrah, varieties that are known for their intense red color, presented the highest values of TA, whereas Garnacha, which has low color intensity, presented the lowest TA value. However, we also detected the uncommon behavior of the Cabernet sauvignon because, although this is considered a highly colored variety, it presented a low value of TA. This is because the field zone where this variety grows is very wet, and this high humidity makes difficult a proper ripening process.

Grape Extracts Spectra. Although the spectral peaks in the MIR frequencies are usually sharp and well resolved, there were two main problems when the phenolic compounds of the studied extracts were analyzed by FT-MIR spectroscopy. The first was that the phenolic compounds considered (anthocyanins, tannins) are chemically very similar and therefore displayed similar MIR absorption characteristics. The second problem was that the contents of water, organic acids, and sugars, which are the major components of the grapes, vary during ripening and all of them absorb in the same MIR region, masking the distinctive MIR vibrations of phenols.[14,28,34−36] In Figure 2 are shown the FT-MIR spectra of 192 grape extracts and, as can be seen, the water and ethanol (from the extraction step)
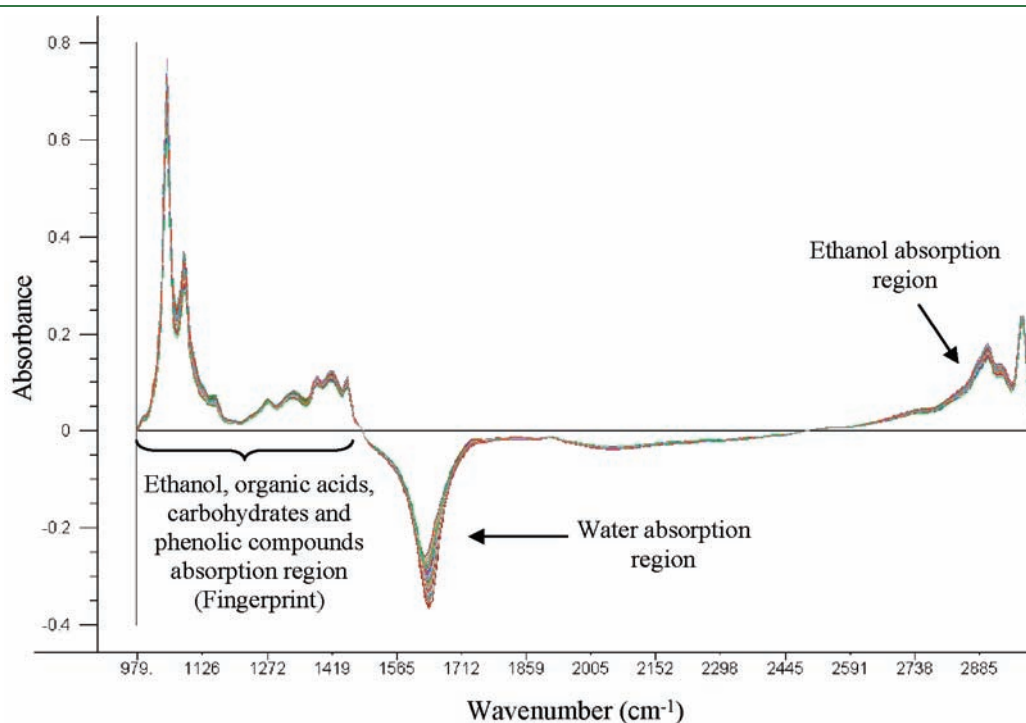

Figure 2. FT-MIR raw spectra of the 192 extracts (979−2989 cm^{-1}) with the three dominant absorption zones pointed out.
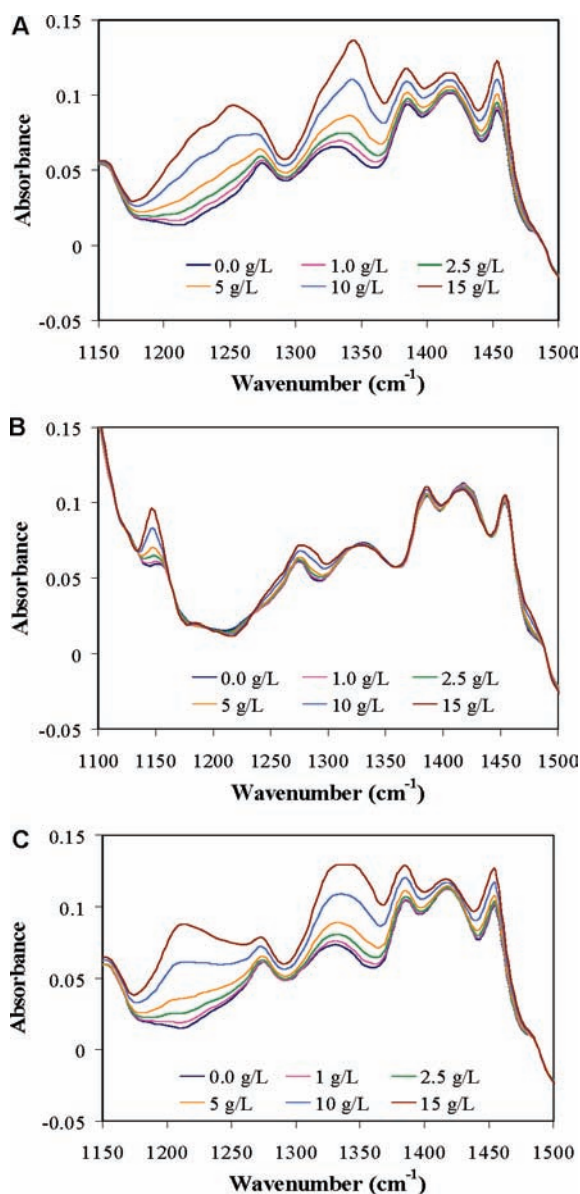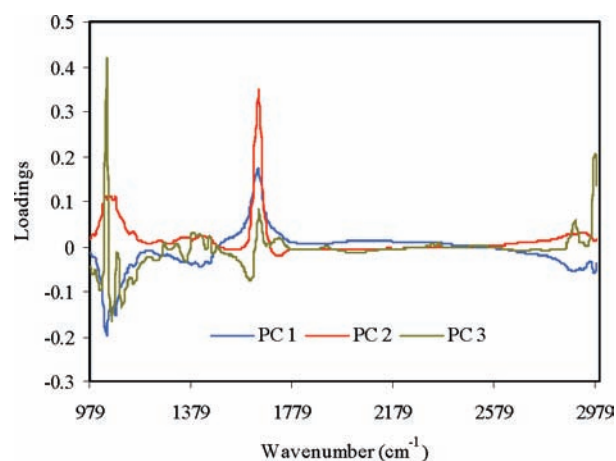
Figure 4. Loadings for the three first PCs on the full range (979–2989 cm^{-1}) corresponding to the 192 sample spectra.

Taking into account this great number of absorption bands, it was not easy to know which were the spectral regions where the phenolic compounds absorbed. Therefore, we investigated the spectral response of four different phenolic compounds by adding increasing amounts of these compounds in a real sample (grape extract). These compounds were gallic acid, (+)-catechin, malvidin-3-glucoside, and tannic acid.

These additions revealed particularly evident variations in the spectral region from 1168 to 1457 cm^{-1} for gallic and tannic acids and in the spectral regions from 1133 to 1160 cm^{-1}, from 1238 to 1322 cm^{-1}, and from 1373 to 1457 cm^{-1} for (+)-catechin (Figure 3). When malvidin-3-glucoside was added to determine the region where anthocyanins absorb, no variations in the spectra were detected due to the low concentration of the commercial standards of these compounds.

Therefore, from these enrichment experiment results, we may conclude that, for phenolic compounds calibration purposes, the best spectral region ranged between 1133 and 1457 cm^{-1}, whereas the remaining wavenumbers (from 979 to 1129 cm^{-1}, from 1457 to 1477 cm^{-1}, and from 2649 to 2989 cm^{-1}) present small and not distinguishing signals of grape phenolic compounds.

Principal Components Analysis. To carry out a study of the information contained in the data set, we applied a PCA considering the response obtained when working with the full range (979–2989 cm^{-1}). The results showed that three principal components (PCs) explained 97.90% of the spectral variation of the samples (first PC, 60.39%; second PC, 35.05%; third PC, 2.46%). Figure 4 shows the loadings plot, which determines the influence of each wavenumber on the variance along this region and, as expected, the three main zones with high influences in the variance of the spectra samples coincided with the three regions where the signal is higher: 979–1477, 1500–1740, and 2869–2989 cm^{-1}. Because the second region corresponds to the water signal and the third region is related to the ethanol contents, it can be concluded that the 979–1477 cm^{-1} region contains almost all of the information that characterizes the samples. Indeed, a PCA of this region showed that the first PC explained 93.77% of variance and that the three first PCs explained 99.41% of variance. Therefore, we considered this region as very useful for calibration purposes and called it the *fingerprint* region. However, taking into account the results of enrichment experiments and also considering that inside this wavenumber range could be observed several sharp absorption bands, it was reasonable

Figure 3. Spectral responses of different concentrations of gallic acid (A), (+)-catechin (B), and tannic acid (C) added to a grape extract sample.

absorption peaks dominate the spectrum. For ethanol, the most intense bands are located at 1045 and 1083 cm^{-1} due to the contribution of C–O stretch (oxygen belonging to a primary alcohol). Other less intense bands located around 2850–2960 cm^{-1} are due to C–H stretch.[36,37] For water, a negative absorption band located in the 1500–1740 cm^{-1} region can be observed. This negative peak is due to the automatic subtraction of the blank that the equipment makes by using distilled water prior to the sample analysis.[28] Related to other compounds present in the samples, on the one hand can be seen an absorption band from 900 to 1100 cm^{-1}, which can be assigned to C–O valence vibrations and C–O–C stretching vibrations of the carbohydrates, including fructose and glucose.[14,36,38] On the other hand, it is well-known that organic acids present an alcohol functional group and, therefore, C–O and the O–H bonds. These different bonds absorb around 1060–1150 cm^{-1} due the C–O stretch (oxygen belonging to a secondary alcohol) and around 1320–1420 cm^{-1} due the O–H bend.[28]

Table 3. Analytical Performance Parameters of the Multivariate Calibration Models Built by PLS Regression Using Different Spectral Regions of Grape Extracts[a]

dependent variable	model	region (cm^{-1})	PLS factors	RMSEC	RMSEC (%)	R^2_{cal}	RMSEP	RMSEP (%)	R^2_{val}	RMSEP$_{cor}$(%)	RPD
TPC	full-range	979−2989	13	56.5	3.5	0.976	72.5	4.5	0.946	4.3	4.3
	fingerprint	979−1477	11	74.7	4.6	0.956	86.9	5.4	0.935	5.3	3.7
	main phenolic region	1133−1457	8	83.4	5.1	0.947	85.7	5.3	0.930	5.2	3.7
	selected region	1168−1457	9	67.3	4.1	0.965	70.6	4.3	0.951	4.1	4.5
TA	full-range	979−2989	15	28.0	3.4	0.988	38.9	4.9	0.959	4.8	4.8
	fingerprint	979−1477	10	47.3	5.8	0.966	51.4	6.5	0.930	6.5	3.6
	main phenolic region	1133−1457	9	48.3	6.0	0.965	51.1	6.4	0.927	6.4	3.5
	selected region	1168−1457	10	41.3	5.0	0.973	46.8	5.9	0.928	5.9	3.5
CT	full-range	979−2989	11	130.6	5.7	0.947	171.6	7.6	0.905	7.5	2.9
	fingerprint	979−1477	10	121.4	5.3	0.961	179.5	8.0	0.900	7.9	2.9
	main phenolic region	1133−1457	10	121.5	5.3	0.957	134.0	5.9	0.934	5.8	3.7
	selected region	1168−1457	9	128.2	5.6	0.953	130.8	5.8	0.937	5.7	3.8

[a] Calibration set n = 106 samples; validation set = 86 samples. TPC, total phenolic compounds in mg of gallic acid kg^{-1}; TA, total anthocyanins in mg of malvidin-3-glucoside kg^{-1}; CT, condensed tannins in mg of (+)-catechin kg^{-1}; RMSEC, root mean square error of calibration; RMSEP, root mean square error of prediction; R^2, coefficient of determination; RPD, residual predictive deviation; RMSEP$_{cor}$, bias-corrected RMSEP.

to consider other smaller intervals in the calibration. After different attempts, the intervals that were shown to contain more information were the 1133−1457 and 1168−1457 cm^{-1} ranges, both coincident with some of the ranges chemically determined.

Quantitative Analysis: PLS Models. When one is working with IR, it is important to determine whether spectra preprocessing is necessary before data analysis. However, because our FT-MIR instrument allows high levels of signal stability and reproducibility to be achieved over time, we decided not to preprocess the spectral data and work with the raw spectral data.

With a quantification purpose, we used the PLS regression method to build four different multivariate calibration models between the FT-MIR spectra (using the four wavenumber ranges selected) and the reference values of the phenolic concentration. The results of each model are reported in Table 3.

As can be seen, in most of cases, the four regions tested provided good calibration models (RMSEC% < 6.0) for the three types of phenolic compounds with accurate prediction (RMSEP % < 8.0 and RPD > 3.0). Nevertheless, it can be also observed that the regions where we found the main phenolic spectral features were, in general, the best correlated with the concentration, providing the best results of the parameters that evaluate the accuracy assessment such as RPD (>3.0) or RMSEP % (<6.0).

Total Phenolic Compounds (TPC) Prediction. To verify the absence of bias between the reference values and the predicted values, we carried out a *joint confidence region* analysis. With this study we could conclude that the results obtained by the four FT-MIR models are reliable because, with a significance level of 0.05, the slope and intercept of the four regression lines obtained were not significantly different from 1 and 0, respectively.

Indeed, when dealing with this phenolic parameter, the RMSEP % values obtained were satisfactory for all regions tested, but the smallest calibration error and also the best R^2_{cal} were obtained when we used the full-range region. However, seeing the relatively high number of factors and also the high difference between the values of RMSEC and RMSEP, we concluded that this model was slightly overfitted.

The comparison of the other three models showed that the 1168−1457 cm^{-1} range provided the best results with both calibration and prediction errors. This behavior suggested that
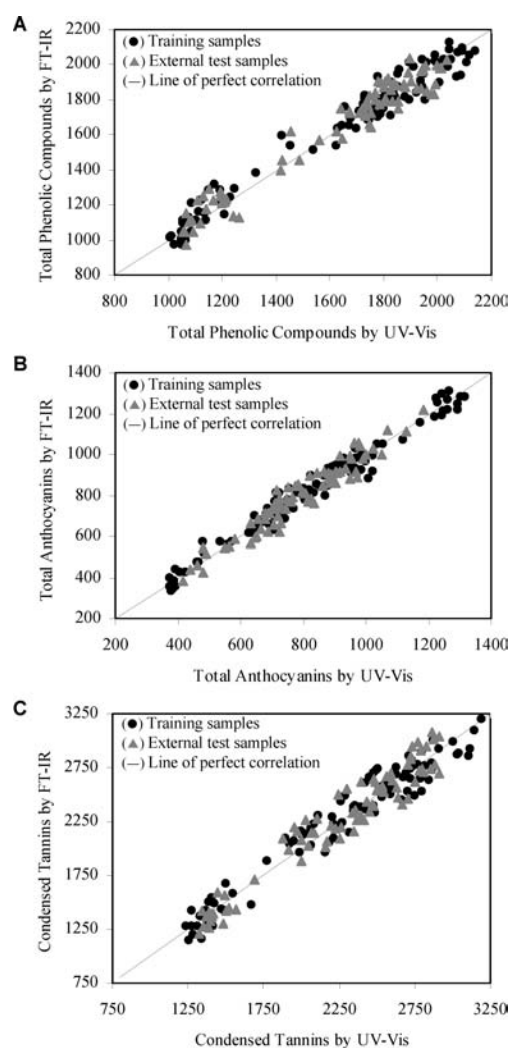

Figure 5. Correlation plot of training and test for the prediction of total phenolic compounds (expressed as mg kg^{-1} of gallic acid) (A), total anthocyanins (expressed as mg kg^{-1} of malvidin-3-glucoside) (B), and condensed tannins (C) (expressed as mg kg^{-1} of catechin) using selected region model (1168−1457 cm^{-1}).

Table 4. Analytical Performance Parameters of the Individual Multivariate Calibration Models Built by PLS Regression for Each Variety and Using the 1168−1457 cm^{-1} Spectral Region[a]

grape variety	dependent variable	samples	PLS factors	RMSEC	RMSEC (%)	R^2_{cal}	RMSECV	RMSECV (%)	R^2_{val}	RPD
Merlot	TPC	27	10	5.0	0.3	0.998	22.7	1.2	0.953	4.7
	TA	29	7	14.6	1.4	0.988	24.2	2.4	0.966	5.5
	CT	21	9	15.8	0.6	0.996	66.7	2.8	0.923	3.7
Tempranillo	TPC	24	8	7.5	0.4	0.992	17.9	1.0	0.954	4.7
	TA	22	7	10.1	1.2	0.992	18.7	2.3	0.971	6.0
	CT	23	6	24.4	0.8	0.989	33.4	1.4	0.979	7.1
Syrah	TPC	24	5	18.1	1.0	0.980	25.6	1.4	0.961	5.2
	TA	30	9	7.7	0.8	0.997	24.2	2.4	0.972	6.1
	CT	22	7	21.0	0.9	0.976	41.7	1.7	0.906	3.3
Cariñena	TPC	20	10	1.8	0.2	0.998	12.7	1.2	0.904	3.3
	TA	26	8	5.1	0.8	0.995	15.4	2.4	0.955	4.8
	CT	20	7	7.3	0.5	0.986	25.0	1.8	0.846	2.6
Garnacha	TPC	20	7	8.1	0.7	0.991	14.9	1.3	0.971	6.0
	TA	20	6	6.1	1.5	0.979	8.5	2.1	0.959	5.1
	CT	22	5	46.8	2.5	0.967	67.2	3.7	0.932	3.9
Cabernet sauvignon	TPC	27	7	13.9	0.8	0.991	23.2	1.4	0.974	6.3
	TA	23	8	7.9	1.0	0.982	16.1	2.1	0.927	3.8
	CT	25	6	45.0	1.8	0.978	73.0	2.9	0.943	4.3

[a] TPC, total phenolic compounds in mg of gallic acid kg^{-1}; TA:, total anthocyanins in mg of malvidin-3-glucoside kg^{-1}; CT, condensed tannins in mg of (+)-catechin kg^{-1}; RMSEC, root mean square error of calibration; RMSEP, root mean square error of prediction; R_2, coefficient of determination; RPD, residual predictive deviation.

this spectral region was more specific for the phenolic compounds than the other regions considered or that the masking effect produced by the interfering substances is less important.[28] Moreover, when the correction of the RMSEP was made, we obtained a RMSEP$_{cor}$ of 4.1%, so this good value, together with the RPD of 4.5, ensures the robustness of the model for the prediction of TPC.

Figure 5A shows graphically the good correlation plot of calibration and validation for the prediction of TPC using the selected 1168−1457 cm^{-1} wavenumber range. Indeed, although from the distribution of the samples shown in Figure 5A, it can be thought that it shows bimodality (due to the fact that some grape varieties presented lower amounts of phenolic compounds than others), it was not necessary to build two calibrations because the validation samples of the external set (i.e., test set) were correctly predicted. Therefore, this behavior demonstrated that there was no overfitting.

Total Anthocyanins (TA) Prediction. For this phenolic parameter the *joint confidence region* analysis showed that there is no bias between the reference and the predicted values, so they are comparables. As can be observed in Table 3, the four wavenumber intervals assayed showed good values for all of the parameters, and these were slightly higher than those of the TPC.

From the results presented in Table 3, we conclude that the best TA prediction was obtained when using the full range because this model presented the smallest error values, the best R^2_{val}, and the highest RPD. However, the difference between RMSEC and RMSEP is the highest, and the number of PLS factors is also higher than those necessary to build the other models. Indeed, the models obtained when using the other regions are also very suitable, especially the one obtained in the selected region 1168−1457 cm^{-1}, as can be seen in Figure 5B.

Because we did not chemically corroborate the spectral features of the anthocyanins and, taking into account the heterogeneous

nature of these pigments given by the presence of glycoside anthocyanins, it is reasonable to think that the relevant spectral information could not be totally coincident with the main phenolic spectral frequencies. Therefore, for these compounds, although a good calibration was obtained in the 1168−1457 cm^{-1} region, it is possible to obtain better calibration and prediction results when more spectral information is considered (i.e., full range). However, the corrected prediction error obtained, RMSEP$_{cor}$, although acceptable whatever the region selected, was always significantly higher than the standard error of the reference method due to the low value of this latest.

Condensed Tannins (CT) Prediction. When evaluating the data for this parameter, we had to remove some samples from the training set because they behaved as outliers. The outliers were identified by their high values of residual y-variance and leverage and subsequently subtracted from the model. The presence of outliers can be attributed, on the one hand, to the fact that the degree of polymerization of these compounds may be very different depending on the grape characteristics and, on the other hand, to the fact that the value of this parameter is strongly dependent on the method used for its determination. Taking into account that the methylcellulose precipitation assay is an indirect method of measurement and that it determines only the phenols with a high size, the accuracy of this method, even suitable, could make more difficult the correspondence with the spectral response. Once these samples were obviated, the models built were satisfactory (Table 3) and the *joint confidence region* analysis for all of the models constructed showed the absence of significant bias. However, we could clearly see that the spectral information obtained chemically helped to improve the model. Thus, whereas the wider regions provided the less accurate prediction (with the highest RMSEP % values and the worst R^2_{val}) and the models constructed in these regions are less robust (RPD < 3.0), the models obtained in the regions in which the enrichment

experiments gave the best results, mainly when the selected region of 1168−1457 cm^{-1} was used. Figure 5C illustrates the good correlation obtained for this parameter in this region. As occurred with TPC correlation, it could be thought that Figure 5C shows bimodality but, because the validation samples of the external set were correctly predicted, we could conclude that there was no overfitting.

Models for Individual Grape Varieties. Provided the good results obtained with the full data set of the samples analyzed and especially in the spectral range of 1168−1457 cm^{-1}, we decided to check the FT-MIR spectroscopy ability to predict the phenolic parameters but working with each grape variety individually.

For this purpose and taking into account the smaller number of samples, the PLS models were constructed using leave-one-out cross-validation. As can be seen in Table 4, for all of the grape varieties the calibration error was very good, with values of <2.5%, and also the calibration lines obtained were very suitable. On the other hand, the low values of the error obtained by cross-validation (<4.0%) also showed the good relationship between the spectra of each variety and the reference values of phenolic compounds. The values of RPD of >3.0 gave an idea of the acceptable robustness of the models built. Only for the CT model of Cariñena did we obtain a not good enough R^2_{val} value (0.846), which coincides with the lowest RPD value. This different behavior can be attributed to the lower range of this parameter values on the calibration samples because the low contents of phenolic compounds is a varietal characteristic of Cariñena grapes. This model only could be applied for rough screening.

From all of these results we conclude that FT-MIR spectrometry combined with multivariate calibration allows a simultaneously fast and accurate determination of TPC, TA, and CT concentrations in grape extracts. The possibility of predicting these grape parameters is an invaluable tool when a large of number of samples needs to be analyzed, as occurs when a phenolic ripening control is necessary.

Moreover, the present work also presents a preliminary attempt to apply the FT-MIR instrument to predict the phenolic composition of specific grape varieties, which is a starting point for the design of specific models according to the requirements of the wineries if a greater number of samples were considered.

■ **AUTHOR INFORMATION**

Corresponding Author
*Phone: +34977558494. Fax: +34977558446. E-mail: montserrat.mestres@urv.cat.

■ **ACKNOWLEDGMENT**

■ **REFERENCES**

(1) Saint-Cricq, N.; Vivas, N.; Glories, Y. Maturation phénolique des raisins rouges. Relation avec la qualité des vins. Comparaison des cépages Merlot et Tempranillo. *Prog. Agric. Vitic.* **1998**, *115*, 306–318.

(2) Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Dubourdieu, D. *Handbook of Enology. Vol. 2. The Chemistry of Wine, Stabilization and Treatments*; Wiley: Chichester, U.K., 1999.

(3) Cheynier, V.; Duenas-Paton, M.; Salas, E.; Maury, C.; Souquet, J. M.; Sarni-Manchado, P.; Fulcrand, H. Structure and properties of wine pigments and tannins. *Am. J. Enol. Vitic.* **2006**, *57*, 298–305.

(4) Romero-Cascales, I.; Ortega-Regules, A.; Lopez-Roca, J. M.; Fernandez-Fernandez, J. I.; Gomez-Plaza, E. Differences in anthocyanin extractability from grapes to wines according to variety. *Am. J. Enol. Vitic.* **2005**, *56*, 212–219.

(5) Rodríguez Montealegre, R.; Romero Peces, R.; Chacón Vozmediano, J. L.; Martínez Gascueña, J.; García Romero, E. Phenolic compounds in skins and seeds of ten grape *Vitis vinifera* varieties grown in a warm climate. *J. Food Compos. Anal.* **2006**, *19*, 687–693.

(6) De Andres-de Prado, R.; Yuste-Rojas, M.; Sort, X.; Andres-Lacueva, C.; Torres, M.; Lamuela-Raventos, R. Effect of soil type on wines produced from *Vitis vinifera* L. cv. Grenache in commercial vineyards. *J. Agric. Food Chem.* **2007**, *55*, 779–786.

(7) Perez-Lamela, C.; Garcia-Falcon, M.; Simal-Gandara, J.; Orriols-Fernandez, I. Influence of grape variety, vine system and enological treatments on the colour stability of young red wines. *Food Chem.* **2007**, *101*, 601–606.

(8) Gomez-Miguez, M.; Heredia, F. Effect of the maceration technique on the relationships between anthocyanin composition and objective color of Syrah wines. *J. Agric. Food Chem.* **2004**, *52*, 5117–5123.

(9) Blouin, J. *Techniques d'Analyses des Moûts et des Vins*; Dujardin-Salleron: Paris, France, 1992.

(10) Mercurio, M. D.; Dambergs, R. G.; Herderich, M. J.; Smith, P. A. High throughput analysis of red wine and grape phenolics-adaptation and validation of methyl cellulose precipitable tannin assay and modified somers color assay to a rapid 96 well plate format. *J. Agric. Food Chem.* **2007**, *55*, 4651–4657.

(11) Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.

(12) Versari, A.; Parpinello, G. P.; Mattioli, A. U.; Galassi, S. Determination of grape quality at harvest using Fourier-transform mid-infrared spectroscopy and multivariate analysis. *Am. J. Enol. Vitic.* **2008**, *59*, 317–322.

(13) Urtubia, A.; Pérez-Correa, J. R.; Meurens, M.; Agosin, E. Monitoring large scale wine fermentations with infrared spectroscopy. *Talanta* **2004**, *64*, 778–784.

(14) Mazarevica, G.; Diewok, J.; Baena, J. R.; Rosenberg, E.; Lendl, B. On-line fermentation monitoring by mid-infrared spectroscopy. *Appl. Spectrosc.* **2004**, *58*, 804–810.

(15) Vonach, R.; Lendl, B.; Kellner, R. High-performance liquid chromatography with real-time Fourier-transform infrared detection for the determination of carbohydrates, alcohols and organic acids in wines. *J. Chromatogr., A* **1998**, *824*, 159–167.

(16) Moreira, J. L.; Santos, L. Analysis of organic acids in wines by Fourier-transform infrared spectroscopy. *Anal. Bioanal. Chem.* **2005**, *382*, 421–425.

(17) Patz, C. D.; Blieke, A.; Ristow, R.; Dietrich, H. Application of FT-MIR spectrometry in wine analysis. *Anal. Chim. Acta* **2004**, *513*, 81–89.

(18) Gorinstein, S.; Weisz, M.; Zemser, M.; Tilis, K.; Stiller, A.; Flam, I.; Gat, Y. Spectroscopic analysis of polyphenols in white wines. *J. Ferm. Bioeng.* **1993**, *75*, 115–120.

(19) Soriano, A.; Pérez-Juan, P. M.; Vicario, A.; González, J. M.; Pérez-Coello, M. S. Determination of anthocyanins in red wine using a newly developed method based on Fourier transform infrared spectroscopy. *Food Chem.* **2007**, *104*, 1295–1303.

(20) Versari, A.; Boulton, R.; Thorngate, J. The prediction of the color components of red wines using FTIR, wine analyses, and the method of partial least squares. *ACS Symp. Ser.* **2004**, *No. 886*, 53–67.

(21) Fernandez, K.; Agosin, E. Quantitative analysis of red wine tannins using Fourier-ransform mid-infrared spectrometry. *J. Agric. Food Chem.* **2007**, *55*, 7294–7300.

(22) Versari, A.; Boulton, R. B.; Parpinello, G. P. Effect of spectral pre-processing methods on the evaluation of the color components of

red wines using Fourier-transform infrared spectrometry. *Ital. J. Food Sci.* **2006**, *18*, 423–431.

(23) Vera, L.; Aceña, L.; Boqué, R.; Guasch, J.; Mestres, M.; Busto, O. Application of an electronic tongue based on FT-MIR to emulate the gustative mouthfeel "tannin amount" in red wines. *Anal. Bioanal Chem.* **2010**, *397*, 3043–3049.

(24) Jensen, J. S.; Egebo, M.; Meyer, A. Identification of spectral regions for the quantification of red wine tannins with Fourier transform mid-infrared spectroscopy. *J. Agric. Food Chem.* **2008**, *56*, 3493–3499.

(25) Janik, L. J.; Cozzolino, D.; Dambergs, R. G.; Cynkar, W. U.; Gishen, M. The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks. *Anal. Chim. Acta* **2007**, *594*, 107–118.

(26) Cozzolino, D.; Cynkar, W. U.; Dambergs, R. G.; Mercurio, M. D.; Smith, P. A. Measurement of condensed tannins and dry matter in red grape homogenates using near infrared spectroscopy and partial least squares. *J. Agric. Food Chem.* **2008**, *56*, 7631–7636.

(27) Fragoso, S.; Mestres, M.; Busto, O.; Guasch, J. Estudio y optimización de un método analítico para la estimación de los parámetros de madurez fenólica de uvas tintas. *Enologos* **2009**, *58*, 38–43.

(28) Moreira, J. L.; Santos, L. Spectroscopic interferences in Fourier transform infrared wine analysis. *Anal. Chim. Acta* **2004**, *513*, 263–268.

(29) Martens, H.; Naes, T. *Multivariate Calibration*; Wiley: Chichester, U.K., 1999.

(30) CAMO. *The Unscrambler User's Guide*, version 9.0; Oslo, Norway, 2003.

(31) Wu, D.; He, Y.; Shi, J.; Feng, S. Exploring near and mid-infrared spectroscopy to predict trace iron and zinc contents in powdered milk. *J. Agric. Food Chem.* **2009**, *57*, 1697–1704.

(32) Sinelli, N.; Spinardi, A.; Di Egidio, V.; Mignani, I.; Casiraghi, E. Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum L.*) by near and mid-infrared spectroscopy. *Postharvest Biol. Technol.* **2008**, *50*, 31–36.

(33) Faber, K.; Kowalski, B. R. Improved prediction error estimates for multivariate calibration by correcting for the masurement error in the reference values. *Appl. Spectrosc.* **1997**, *51*, 660–665.

(34) Boulet, J.; Doco, T.; Roger, J. M. Improvement of calibration models using two successive orthogonal projection methods. Application to quantification of wine mannoproteins. *Chemom. Intell. Lab. Syst.* **2007**, *87*, 295–302.

(35) Boulet, J. C.; Williams, P.; Doco, T. A Fourier transform infrared spectroscopy study of wine polysaccharides. *Carbohydr. Polym.* **2007**, *69*, 79–85.

(36) Armenta, S.; Garrigues, S.; de la Guardia, M.; Rondeau, P. Attenuated total reflection-Fourier transform infrared analysis of the fermentation process of pineapple. *Anal. Chim. Acta* **2005**, *545*, 99–106.

(37) Bevin, C. J.; Fergusson, A. J.; Perry, W. B.; Janik, L. J.; Cozzolino, D. Development of a rapid "fingerprinting" system for wine authenticity by mid-infrared spectroscopy. *J. Agric. Food Chem.* **2006**, *54*, 9713–9718.

(38) Edelmann, A.; Diewok, J.; Schuster, K. C.; Lendl, B. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. *J. Agric. Food Chem.* **2001**, *49*, 1139–1145.