Journal of Chromatography A, 1235 (2012) 68–76

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

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

Multi-wavelength high-performance liquid chromatographic fingerprints and chemometrics to predict the antioxidant activity of *Turnera diffusa* as part of its quality control

J. Ricardo Lucio-Gutiérrez [a], Aurora Garza-Juárez [b], J. Coello [a,*], S. Maspoch [a], M.L. Salazar-Cavazos [b], Ricardo Salazar-Aranda [b], Noemi Waksman de Torres [b]

[a] *Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*
[b] *Departamento de Química Analítica, Facultad de Medicina, Universidad Autónoma de Nuevo León, C.P. 64460, Monterrey, Nuevo León, Mexico*

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form 16 February 2012
Accepted 17 February 2012
Available online 23 February 2012

Keywords:
Turnera diffusa
HPLC-fingerprints
Multi-wavelength
Antioxidant activity
Chemometrics

ABSTRACT

The determination of the antioxidant activity of *Turnera diffusa* using partial least squares regression (PLSR) on chromatographic data is presented. The chromatograms were recorded with a diode array detector and, for each sample, an enhanced fingerprint was constructed by compiling into a single data vector the chromatograms at four wavelengths (216, 238, 254 and 345 nm). The wavelengths were selected from a contour plot, in order to obtain the greater number of peaks at each of the wavelengths. A further pretreatment of the data that included baseline correction, scaling and correlation optimized warping was performed. Optimal values of the parameters used in the warping were found by means of simplex optimization. A PLSR model with four latent variables (LV) explained 52.5% of *X* variance and 98.4% of *Y*, with a root mean square error for cross validation of 6.02. To evaluate its reliability, it was applied to an external prediction set, retrieving a relative standard error for prediction of 7.8%. The study of the most important variables for the regression indicated the chromatographic peaks related to antioxidant activity at the used wavelengths.

1. Introduction

Increasing interest in the use of plant-based products is leading to a fast-growing market for herbal pharmaceuticals, dietary supplements, nutraceuticals, and functional foods [1]. This major expansion in the use of herbal medicines raises many concerns about their quality control, as it could lead to massive and unsupervised collection without any consideration of the quality of the raw materials [2]. Moreover, assuring the quality of a herbal medicine is a major challenge in the phytopharmaceutical and food industries, because the chemical content of herbs varies greatly according to a wide range of factors such as species variation, growth location, climate, harvesting season, storage conditions and processing [3,4]. In order to cope with the problems related with this chemical complexity, the researchers adopted the herbal fingerprinting approach, which has been accepted by the World Health Organization (WHO), the US Food and Drug Administration (FDA) and European Medicines Agency (EMA), among other organizations, as a strategy for the assessment of herbal medicines [5–9].

The chromatographic fingerprint of a plant material is a chromatogram obtained by a defined procedure where as many compounds as possible are separated and that represents the chemical characteristics of the herb; usually, samples with similar fingerprints have similar properties [5,10]. Thus, a chromatographic fingerprint could be used to determine the identity, authenticity, batch-to-batch consistency of the herbal medicine and it is also useful to overcome the limitations when using few marker compounds [11,12]. Frequently, one or two markers, principally the pharmacologically active constituents of a plant material, are employed for quality control of herbal medicines [12]. However, this type of analysis may not offer a complete characterization of these products, because their therapeutic effects could originate from many components. Moreover, synergic or antagonist interactions are also ignored in this type of analysis; thus, it is not feasible to analyze each compound individually [13]. Among the chromatographic methods for fingerprinting herbal drugs, high performance liquid chromatography hyphenated to diode array detector (HPLC-DAD) is still the most popular [10,14]. Besides studying the identities of herbs, some publications explore the usefulness of fingerprints collected by HPLC-DAD as a tool

* Corresponding author. Tel.: +34 93 581 2122; fax: +34 93 581 2357.
E-mail addresses: jrluciog@yahoo.com (J.R. Lucio-Gutiérrez), jordi.coello@uab.es (J. Coello).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2012.02.042

to predict the therapeutic properties of medicinal plants. For example, Dumarey et al. [5] and Van Nederkassel [15] showed a correlation between the chromatographic fingerprint (at 280 nm) of green tea samples and their antioxidant capacity, as a measure of protective effects caused by the tea. However, in the correlation of the fingerprints with a possible biological activity, chemists usually use a single chromatogram from a fixed wavelength; so a lot of information contained at other wavelengths may be discarded, neglecting chromatographic peaks that are useful to characterize the properties of a medicinal herb.

Recently, a strategy that involves the use of several wavelengths for the construction of enhanced herbal fingerprints has been used for purposes of classification and counterfeit detection to *Valeriana officinalis* [16]. We have used the enhanced fingerprints strategy in this work, to improve the prediction of the biological activity of herbal medicines as part of their whole quality control. In order to show the usefulness of this multi-wavelength approach, we have applied the procedure to *Turnera diffusa* chromatograms. This plant was selected because organizations (EMA, FDA, etc.) have yet to report a definitive analytical methodology for its quality control and because it was shown that many of the products containing *T. diffusa* in the Mexican market are inconsistent, according to reports where fingerprints obtained by TLC and HPLC-DAD were applied [2,17,18]. A previous study about correlating the chromatographic fingerprints of *T. diffusa* (at 254 nm) and its antioxidant activity using PLSR was performed by Garza-Juárez et al. [13]; however, the limitations when using a single wavelength were evident in the results of the mathematical model and the predictions, obtaining a r^2 from the cross-validation of 0.80, a bias with a slight tendency to negative values and poor performance of the model when predicting an external sample set. Furthermore, the retention time shifts correction was performed manually and therefore the alignment preprocessing was time consuming. On the other hand, the proposed multi-wavelength strategy should improve the performance of the PLSR model and make it possible to obtain more robust predictions. Moreover, correlation optimized warping (COW) algorithm is applied to facilitate the alignment preprocessing.

T. diffusa Wild. Ex Schult, also known as damiana, is a small shrub belonging to the family *Turneraceae*; it grows to a height of 1–2 m and bears 10–25 cm long aromatic, serrated leaves. Small yellow flowers bloom in summer, and are followed by small fruit with a sweet smell and a fig-like flavor. The medicinal part of the plant is its leaves which are harvested, according to ethnopharmacy practices, during the flowering season [18,19]. The damiana shrub can be found throughout Mexico, Central America, the Caribbean islands and parts of South America. It has several traditional uses as an aphrodisiac, for hepatic ailments, depression, anxiety, neurosis, as expectorant, stimulant and tonic; it is also used to flavor desserts, beverages, candies, etc. [20]. In addition, has been reported that the aerial part (stems and leaves) showed good antioxidant activity, similar to that exhibited by quercetin, used as a control [21]. Antioxidants are known to reduce the risks of certain types of cancer and many chronic degenerative diseases; so, there is an especially fast-growing interest in the search for naturally occurring antioxidants, in order to obtain diverse products that are at most used as dietary supplements [22,23]. We consider that the attributed therapeutic effects of damiana are related, at least to some extent, with its antioxidant activity; thus, it can be a measure for the effectiveness and for the quality of the herb. Multi-wavelength chromatographic fingerprints are a good choice to predict the antioxidant activity of damiana, and other medicinal herbs, because they represent most of the constituents of herbal products and because it is advantageous to handle these enhanced fingerprints using several chemometric methods.

2. Materials and methods

2.1. Chemicals and samples

HPLC-grade methanol (MeOH) was purchased from Fisher Scientific (Fair Lawn, NJ, USA). Deionized water was obtained from Laboratorios Monterrey S.A. de C.V. (Monterrey, NL, Mexico). Trifluoroacetic acid (TFA), reagent grade MeOH and ethanol (EtOH) were purchased from Fermont (Monterrey, NL, Mexico). The 2,2-diphenyl-1-picryl-hydrazyl (DPPH) and the 2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxychromen-4-one (Quercetin), used to obtain the scavenging activity by spectrophotometry, were purchased from Sigma (Monterrey, NL, Mexico). In order to include as many sources of sample variability as possible, regardless of the specific geographical origin, forty samples of *T. diffusa* were collected in different regions of Mexico between December 2005 and January 2009; the samples were authenticated and voucher specimens deposited in the herbarium of the Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León. Damiana samples were stored protected from bright light and moisture at laboratory temperature (27 °C approx.) and allowed to dry.

2.2. Preparation of the extracts

The aerial parts (leaves and stems) of the dried plants were ground and then passed through a 40 mesh sieve. Each sample powder (1.00 g) was accurately weighed and extracted three times using 5 mL 9:1 (v/v) ethanol–water solution every time, at 27 °C, by vortex mixing for 3 min and respective solutions were combined. The final extract was filtered through a Whatman No. 40 filter paper and evaporated to dryness under reduced pressure by rotary evaporation at 37 ± 2 °C. Before use, 15 mg of the sample extract was taken and dissolved in 1 mL of MeOH and filtered through a 0.45 μm nylon acrodisc (Waters Corporation).

2.3. Instrumentation and analytical procedures

HPLC analyses were performed with a Waters 2695 Alliance system, equipped with a 2996 UV–Vis diode array detector, an autosampler, a thermostated column compartment, a vacuum degasser and a PC with Empower software. The radical scavenging activity of damiana extracts was obtained using a UV–Vis Beckman DU-7500 spectrophotometer.

2.3.1. HPLC-DAD procedure

The separation was achieved using an AccQ Tag C18 column; 3.9 mm × 150 mm, 4 μm particle size (from Waters Corporation) operated at 30 °C. The mobile phase consisted of (A) 0.1% TFA in water and (B) MeOH. Before use, the mobile phase constituents were degassed and filtered through a 0.45 μm nylon filter (Waters Corporation, Milford, MA, USA). The gradient elution sequence was used as follows: initial 30% of B; then, the percentage of B was increased from 30 to 70% over 30 min and held constant for 5 min. Each run was followed by an equilibration period of 10 min with the initial conditions (70% A, 30% B). Before the first injection, the column was preconditioned for 1 h with the initial mobile phase. The flow rate was kept at 0.4 mL min^{-1} and the injection volume was 10 μL. The DAD was set at 254 nm to follow the elution profile, UV spectra were collected in the range 210–400 nm at 1.2 nm steps and the sampling rate of the detector was 0.01667 min.

2.3.2. Determination of the scavenging activity of DPPH free radical using spectrophotometry

An assay of DPPH scavenging activity using spectrophotometry was conducted according to Leu et al. [24], with some modifications. First, the extract was redissolved in ethanol (1 mg mL^{-1}), and

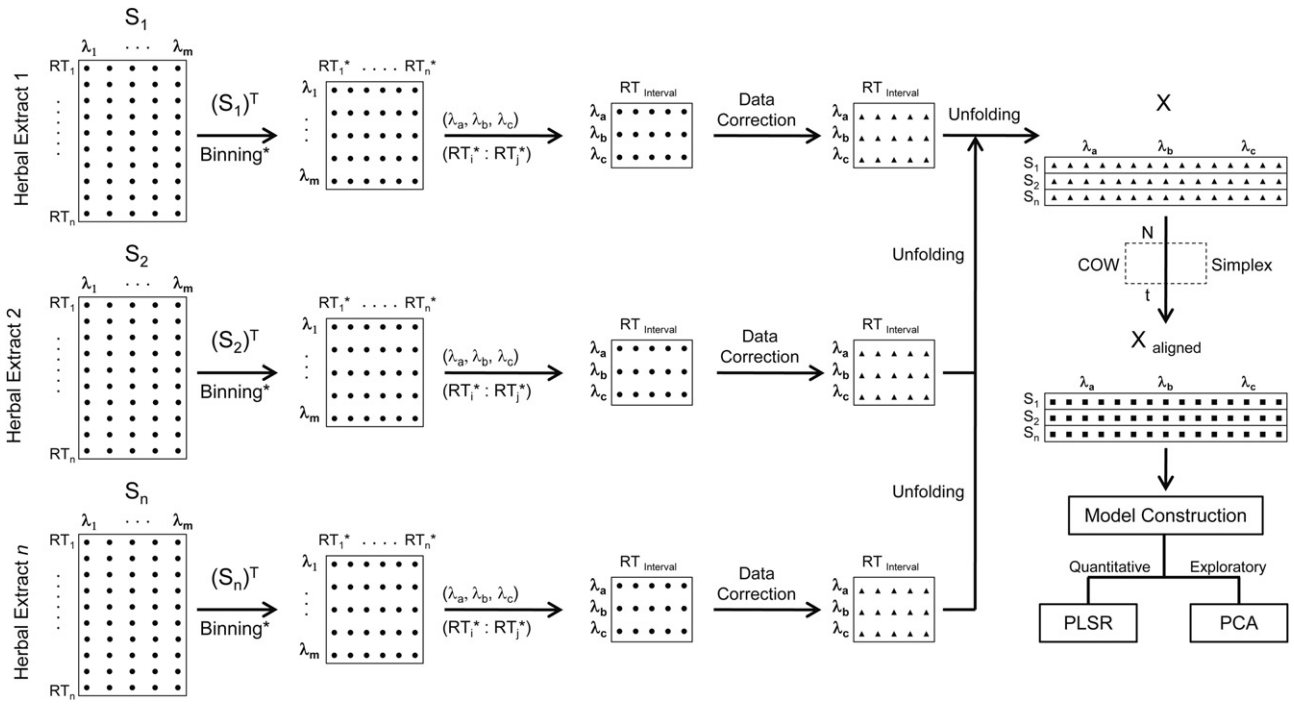

Fig. 1. Schema of the construction of the enhanced fingerprints. $S_1, S_2, \ldots, S_n$; represent original two-dimensional chromatographic data (retention time [RT] × wavelengths) from different herbal extracts. X is the newly constructed data matrix of enhanced fingerprints. COW, N and t are the acronyms and the symbols for correlation optimized warping, segment length and slack parameter, respectively. $X_{aligned}$, is the final data matrix to be used in the construction of exploratory data analysis (PCA) and quantitative (PLSR) models.

different concentrations (0.2–200 μg mL^{-1}) of each extract were used. In a total volume of 1 mL, the assay mixture contained 500 μL of the extract and 500 μL of DPPH (125 μM in ethanol). The assay mixture was shaken and allowed to stand at room temperature in darkness for 30 min. The absorbance was then measured at 517 nm. Quercetin was used as a positive control. The capacity to scavenge the DPPH radical was calculated as follows:

$$\text{Radical scavenging activity } (\%) = \left(\frac{[A-B]}{A}\right) \times 100 \quad (1)$$

where A is the absorbance of the negative control (DPPH plus ethanol), and B is the absorbance of the sample (DPPH, ethanol plus sample). Concentration was plotted vs. percentage of scavenging and the effective concentration of the extract that scavenged 50% of DPPH radicals (EC_{50}) was calculated by interpolation; so, the activity was expressed as EC_{50}. The antioxidant activities of the samples and the respective chromatograms were obtained simultaneously to avoid possible variations in the results from the DPPH test, due to the storage time of the extracts.

2.4. Data analysis

The used data resulted from a study performed by Garza-Juárez et al. [13]. HPLC-DAD data of forty damiana samples were exported form Empower software in ASCII format and imported to MATLAB 7.9 (The MathWorks, Inc); which was used for handling chromatographic data matrices and to perform their scaling. Data binning (co-adding), baseline correction, unfolding and the construction of mathematical models were performed using PLS-Toolbox 6.2 (Eigenvector Research). The selection of reference chromatogram used in the alignment, the simplex optimization for establishing segment length and slack size and correlation optimized warping (COW) alignment preprocessing were performed using the codes for Matlab developed by Skov [25] and by Tomasi [26], which are freely available from http://www.models.kvl.dk/.

2.4.1. Construction of the multi-wavelength fingerprints

Fig. 1 depicts a general procedure of the approach used for the construction of enhanced herbal fingerprints. A data matrix of a given sample (retention time × wavelengths) is obtained by taking the chromatograms in a range of wavelengths; then, the matrix transpose is calculated in order to subsequently apply the row-wise algorithms to the chromatograms, e.g. data co-adding (binning). Next, the strategy comprises of the selection of both those wavelengths where the chemical constituents of the herbal extract exhibit maximum absorbance, simultaneously presenting a larger number of peaks, and the retention time intervals that contain informative signals and prevent the inclusion of noise. Then, the retrieved chromatographic profiles are subjected to pretreatments to minimize unwanted chromatographic variations, e.g. background correction, and they are reshaped into a single data vector, which is the final fingerprint of the sample. The enhanced fingerprints obtained from all considered samples are compiled into a new chromatographic data matrix that, after alignment if needed, is used to construct exploratory and regression models by means of Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR). Since with this strategy it is possible to include as many wavelengths as selected by the analyst, most of the relevant chromatographic information would be considered in the construction of mathematical models and therefore, this approach could attain better performance than that obtained using a single wavelength; especially in the quantitative results from the PLSR when predicting a specific biological activity.

2.4.2. Enhanced fingerprints alignment by COW

Before the construction of the final matrix of enhanced fingerprints that will be used in antioxidant activity prediction, the COW algorithm was applied to correct the retention time shifts among the different samples. Using the maximum cumulative product of correlation coefficients, a reference data vector was selected from the enhanced fingerprints of authenticated damiana samples. All

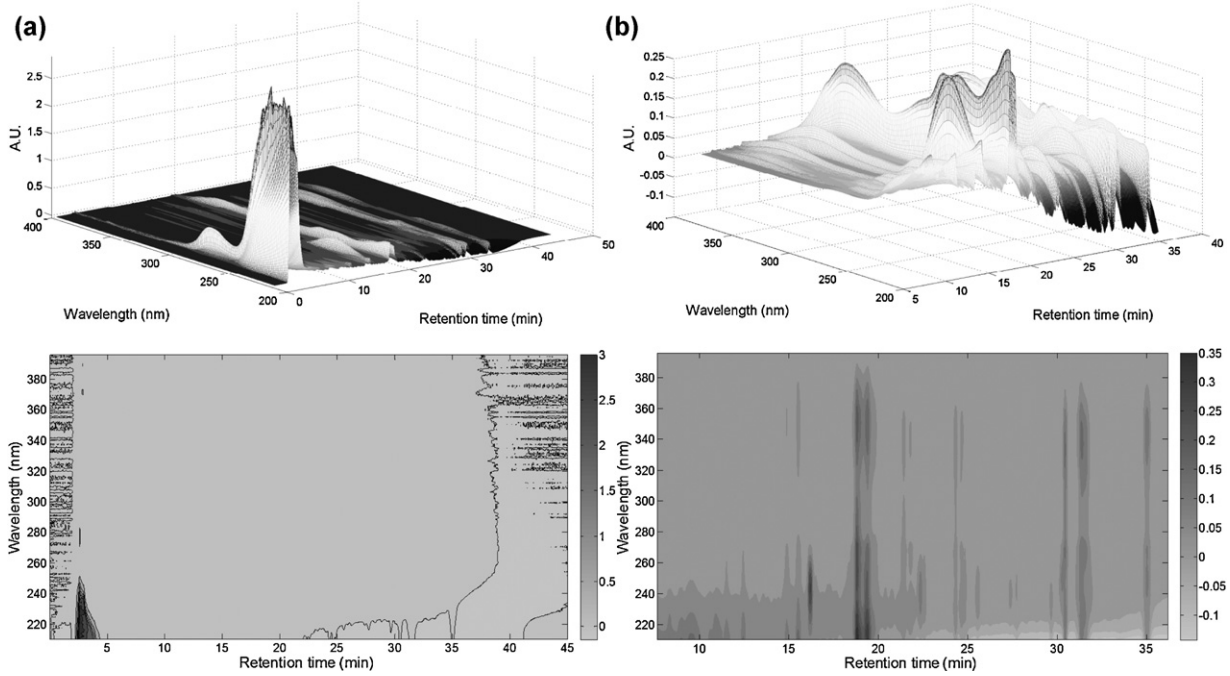

Fig. 2. Above, a 3D chromatogram plot (min × nm × absorbance [A.U.]) from a damiana sample and its respective contour plot below; (a) complete chromatogram and (b) after selection of the retention time interval.

the other samples were aligned with regard to this reference. Segment length and slack size values were found by means of a simplex design, performed within an optimization space defined as follows: segment from 45 to 135 and slack from 4 to 40. These numbers were chosen according to the observed peak widths and shifts on the chromatograms. The number of starts in the grid search was set to 5, the maximum number of optimization steps to 50 and the fraction of maximal deviation from center in COW alignment to 15%. The theory for the algorithms involved in reference selection and COW optimization can be consulted in [25].

2.4.3. Principal component analysis

PCA is a well known bilinear modeling method which gives an interpretable overview of the main information in a multidimensional data table by extracting and displaying the existing systematic variation. The decomposition of the chromatographic data table is performed according to the following equation:

$$X = TP^T + E \quad (2)$$

where X is the final matrix of enhanced fingerprints, T is the scores matrix with as many rows as the original data, P^T corresponds to the loadings matrix transpose with as many columns as the original data and E is an error matrix [27]. The number of columns in the scores matrix corresponds to the principal components (PCs) that are calculated to describe the variance of the data. So, the information carried by the original variables is compressed into a smaller number of uncorrelated variables; i.e., the principal components. By plotting the scores, we are able to detect and interpret sample patterns, groupings, similarities and differences, and to discover outliers. In addition, PCA was carried out to verify the usefulness of the warping performed by evaluating the model results before and after conducting the alignment.

2.4.4. Partial least squares regression

PLSR is used to find the inner relationship between independent variables (X) and dependent variables (Y), which are simultaneously modeled by taking into account not only X variance, but the covariance between X and Y [28]. In our case, the X matrix is composed of the enhanced fingerprints and the Y vector is constructed with the reference values of antioxidant activity (EC_{50}) obtained by the DPPH assay. Then, X and Y are decomposed in a product of another two matrices of scores and loadings; as described by the following equations:

$$X = TP^T + E \quad (3)$$

$$Y = UQ^T + F \quad (4)$$

where TP^T approximates to the chromatographic data and UQ^T to the true Y values; notice that the relationship between T and U scores is a summary of the relationship between X and Y. The terms E and F from the equations are error matrices. Hence, the PLS algorithm attempts to find factors (called Latent Variables) that maximize the amount of variation explained in X that is relevant for predicting Y; i.e., capture variance and achieve correlation [29]. Cross-validation (CV) was used when PCA and PLSR calibration models were developed. The optimum number of factors was determined by the minimum value of predicted residual error sum of squares (PRESS) criterion [30]. Statistics calculated for the calibration model included root mean square error of cross validation (RMSECV) and determination coefficient r-squared.

3. Results and discussion

3.1. Performance of DPPH assay and HPLC method

The percentage reduction of DPPH radical exhibited by the different concentrations of a given sample was determined and subsequently its EC_{50} was calculated; each sample measurement was made in triplicate, obtaining RSD values between 0.1 and 18.1%. The precision of the chromatographic method was evaluated considering relative retention times, relative peak heights and relative areas of 12 peaks common to all the chromatograms, which span over the whole chromatographic region, using the methodology reported in [18]. The results expressed as RSD were between 0.1 and 2.8% for retention times, 2.3 and 12.0% for the relative areas,

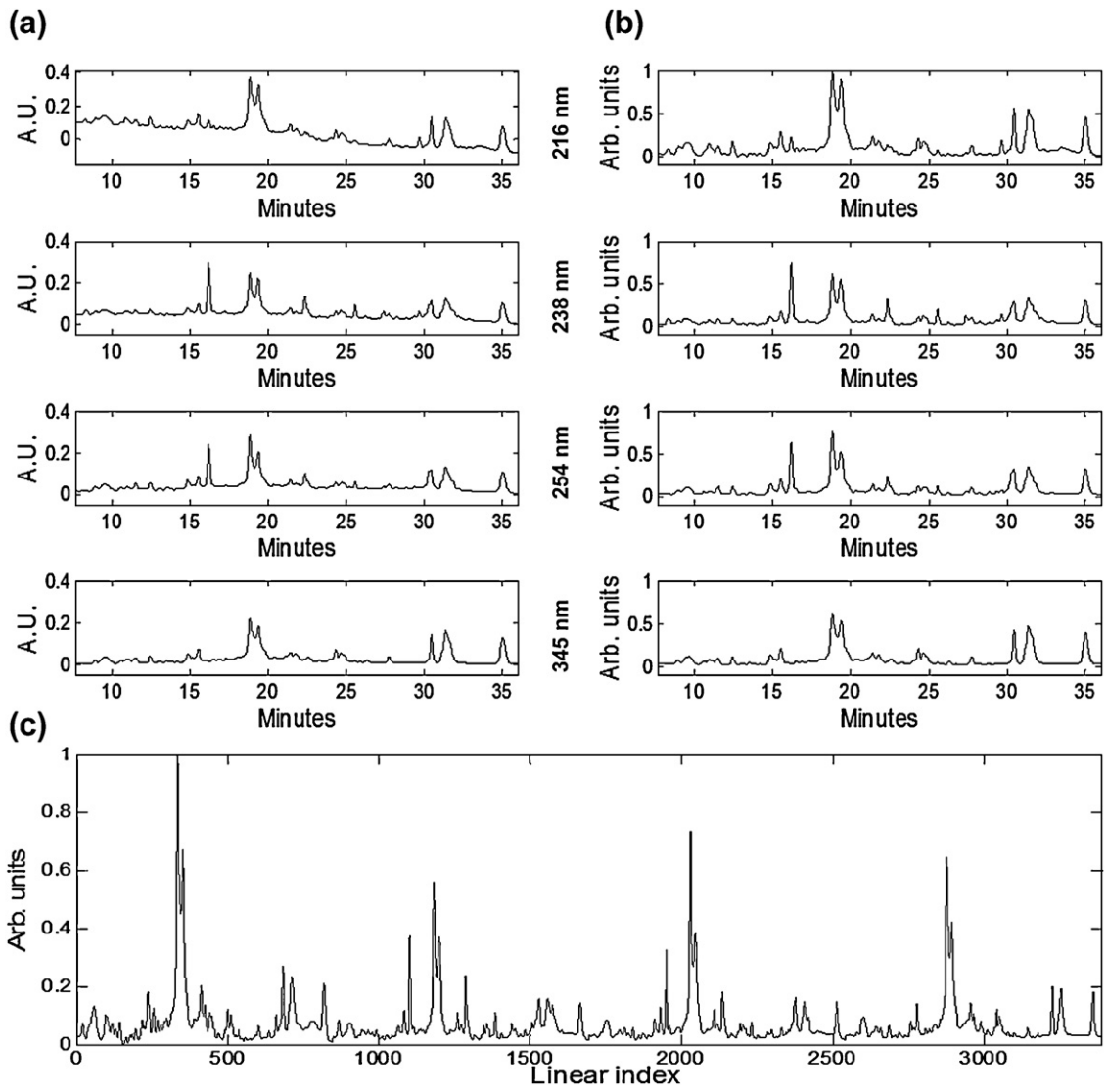

Fig. 3. Chromatograms from a *T. diffusa* sample at wavelengths used in enhanced fingerprint construction; (a) before data pretreatment and (b) after data binning, baseline correction and scaling. A.U. corresponds to absorbance units and Arb. units means arbitrary units. (c) Typical appearance of an enhanced fingerprint from *T. diffusa*.

and 2.6 and 16.0% for the relative heights. Thus, all results indicated acceptable performance of the analytical procedures used.

3.2. Enhanced fingerprints construction and alignment

The chromatograms were very complex, showing about 46 peaks and it is evident that some of them are not fully resolved. For each sample, a chromatographic data matrix with dimensions 2700 × 158 (retention time × wavelengths) was obtained. Fig. 2a shows a raw 3D chromatogram plot from a damiana sample with its respective contour plot. Using the contour plot of the original chromatographic data, it was difficult to obtain information because of the high intensity of retention times from 0 to 7.66 min, a broad and noisy signal that corresponds to the elution front; which contains several non-resolved compounds, and therefore it was removed from further calculations. Also, retention times above 35.86 min were removed because there was no signal. In this way, the better resolved and more stable peaks in terms of area were in the selected retention time interval and the inclusion of noise or non-informative signals were avoided in the mathematical models. The dimensions of the matrix are now (1692 × 158). Fig. 2b shows the chromatogram in the selected retention time interval; now, from the contour plot it is possible to observe that at different wavelengths there are different peaks of different intensities. Most of the chemical constituents detected in extracts of damiana showed the largest responses at four wavelengths: 216, 238, 254 and 345 nm; consequently, the chromatograms from these wavelengths were used to construct the enhanced fingerprint of each sample. Thus, a greater number of peaks of the highest intensity were included and wavelengths with redundant information were excluded; notice that not all peaks are present at all the wavelengths, see Fig. 3a. Next, after performing the data matrix transpose, the retention time dimension was co-added, i.e. adjacent variables were combined, using the mean value from groups of 2 variables and matrix dimensions were (158 × 847 [wavelengths × retention time]); then, the four working wavelengths were retrieved and the final matrix dimensions were (4 × 847). Fig. 3a shows the variations presented by the chromatograms from a *T. diffusa* sample at the selected wavelengths used to construct the enhanced fingerprints. The most evident of these variations was the change in baseline slope; it was also evident in the 3D chromatogram from Fig. 2b. There was a slight tendency towards negative values at 238, 254 and 345 nm and a larger tendency to negative values at 216 nm. In such conditions, it was not viable to unfold the chromatographic data matrix

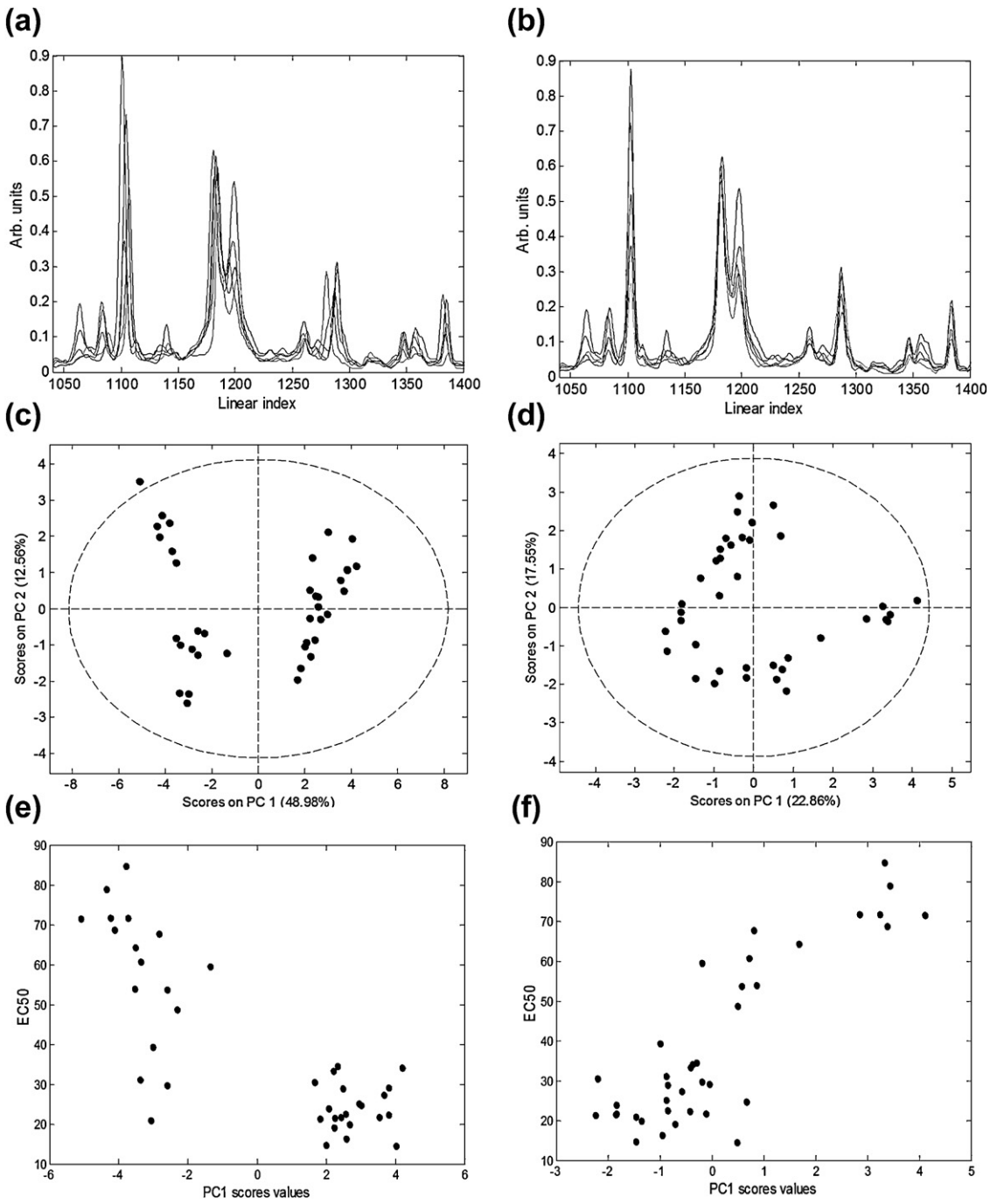

Fig. 4. (a) Section from several enhanced fingerprints before correlation optimized warping and (b) the same samples after alignment. (c) and (d) Score plots of principal component 1 (PC1) vs. PC2 from an overall PCA of enhanced fingerprints before and after alignment, respectively. (e) Plot of the scores from PC1 vs. antioxidant activity values (EC_{50}) for each sample before warping and (f) after warping.

to construct the enhanced fingerprint vector; so, the data had to be corrected. Background was eliminated by means of asymmetric weighted least squares method using a cubic polynomial, which subtracts a baseline from a chromatogram with the constraint that residuals below zero be weighted more heavily than those above zero; details about the method can be found elsewhere [31]. Furthermore, the variability in scaling among the damiana samples was minimized applying a range scale of each variable such that the highest value of the background corrected data vectors was +1 and the lowest 0. Fig. 3b shows the results of applying the pretreatments to vectors of the chromatographic data matrix. The data binning kept the chromatographic information, i.e. did not affect the chromatogram's features, and reduced the number of data points in the retention time dimension of the matrix. The baseline correction successfully removed the tendencies exhibited by the background of the chromatograms; and scaling each enhanced fingerprint to unit, maintained intensity ratios among the four wavelengths used to build the enhanced fingerprint and avoided discontinuities in absorbance values when the four wavelengths were unfolded. So, the corrected data matrix was unfolded along the retention time dimension to generate a single data vector with dimensions of 1×3388, which is the enhanced fingerprint of the sample; Fig. 3c. As a result of the matrix unfolding, the retention time dimension was lost and had to be replaced with a linear index; nonetheless,

the units of the linear index can be translated into retention time units. As expected, the enhanced fingerprint presented more sets of peaks than when exhibited using a single wavelength. Finally, data vectors of all samples were ordered into a matrix for subsequent alignment and construction of the models.

The retention time shift correction was the final data pretreatment to be performed; input parameters were the enhanced fingerprint used as a reference, the segment length and a slack size. Simplex optimization to find the working parameters for the warping retrieved a segment length of 45 and a slack size of 23; Fig. 4a and b presents the original and the aligned enhanced fingerprints of several samples using these warping parameters. The retention time shifts were properly corrected and peaks and features were not distorted by the use of the COW algorithm. The scores plot from a PCA performed before the alignment (Fig. 4c) shows two main groups of samples separated by the PC1; the group in the negative score values of PC1 is also split into two subgroups. Analysis of the loadings from this model suggested that the observed score plot is mainly due to retention time shifts, since a loadings plot (not shown) presented a fuzzy aspect and depicted shapes similar to those obtained when calculating a derivative. Fig. 4d is the scores plot obtained after performing the retention time shifts correction;

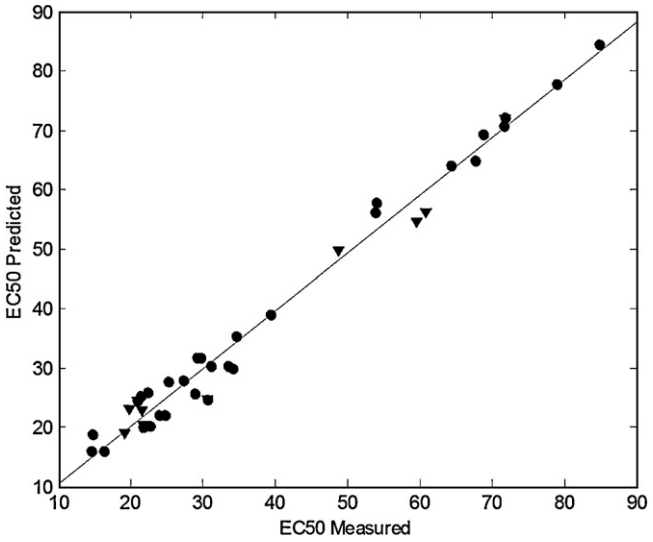

Fig. 5. (●) Calibration; (▼) prediction. Measured vs. Predicted plot from the antioxidant activity (EC_{50}) of each sample; regression line is plotted.

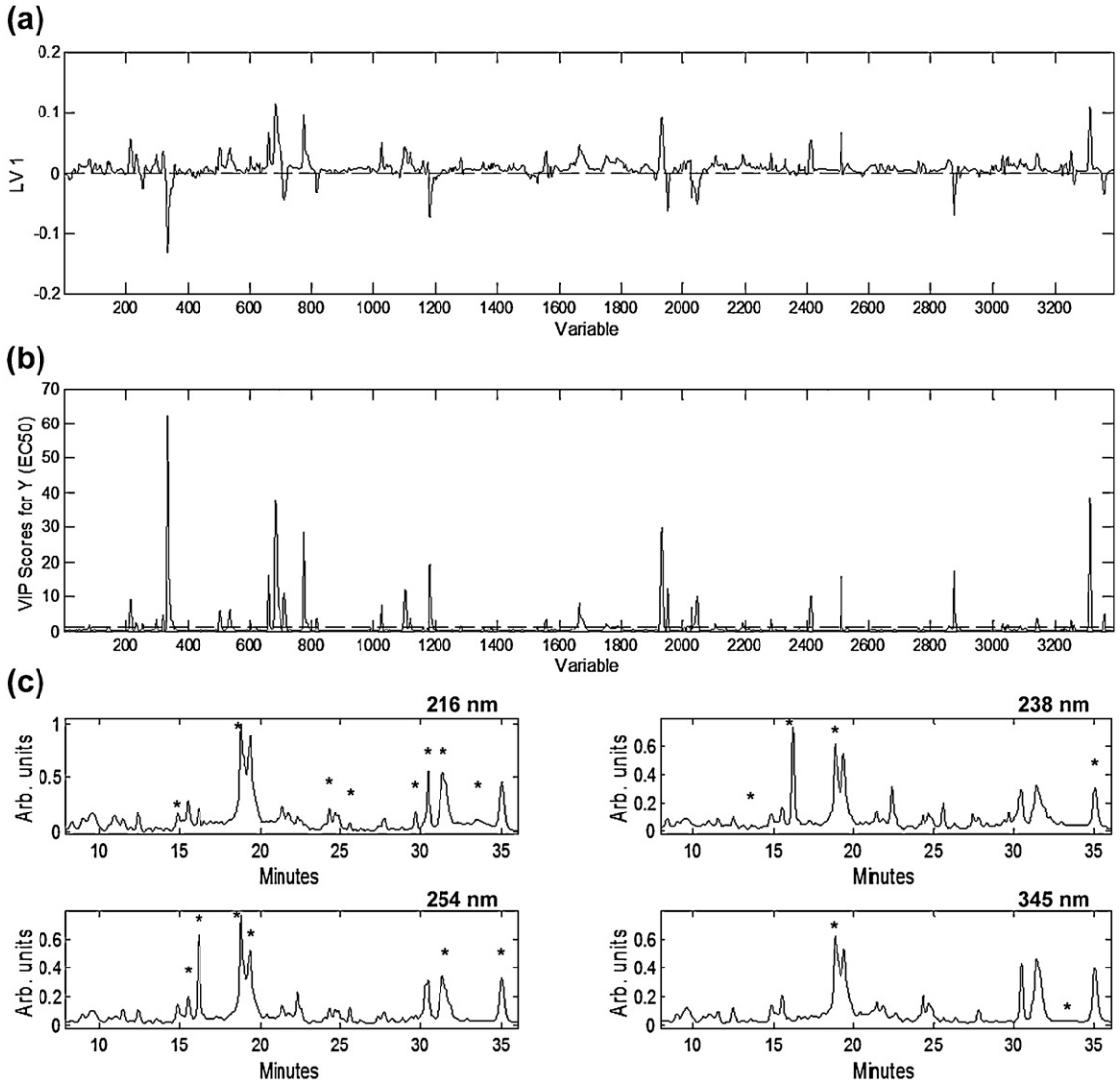

Fig. 6. (a) Loadings plot on LV1 from chromatographic variables used. (b) Variable importance in projection (VIP) scores for EC_{50}; variables with values above the horizontal dashed line being considered important for the model. (c) Indication of the relevant peaks for the prediction of the EC_{50} value, marked by asterisks, on chromatograms from an enhanced fingerprint of a *T. diffusa* sample.

the split of the two subgroups is no longer present and the loadings plot more resembled a chromatogram, suggesting that retention time shifts were corrected successfully. In order to find out whether there is some kind of relationship between the principal components and the antioxidant activity, the scores from each calculated principal component of the PCA vs. the antioxidant activity were plotted. Projection of the samples on the PC1 axis is related with EC_{50} values; when plotting the scores from PC1 vs. the antioxidant activity, after warping, a linear relationship was evident, see Fig. 4f. However, the same plot before alignment does not show any apparent relationship, Fig. 4e. No relationship was obvious between other principal components and EC_{50} values.

3.3. Antioxidant activity (EC_{50}) prediction by PLSR

A PLSR model to relate antioxidant activity and chromatographic data was constructed with the enhanced fingerprints of 30 samples; so, an X matrix of dimensions (30×3388) and its respective Y vector of activities were used. The X data were mean centered and a cross-validation was performed, applying the venetian blinds method with 6 data splits [32]. A model with four latent variables (LV) was chosen according to the minimum PRESS criterion, achieving an explained variance of 52.5% for X variables and 98.4% for Y; a RMSECV of 6.02 and r^2 from CV of 0.928 were obtained. In any case, outliers were not detected during model construction. To evaluate the performance of the model, 10 samples not employed for calibration were used as an external prediction set; the values of these samples span over the calibration range of antioxidant activity. Fig. 5 shows the EC_{50} measured vs. EC_{50} predicted plot that includes calibration and prediction sets, with a slope of 0.972 ($s = 0.022$) and an intercept of 0.88 ($s = 0.94$), indicating that predicted values are coincident with the measured ones. Relative standard error of prediction (RSEP) was calculated as:

$$\%\mathrm{RSEP} = \left(\frac{\sum [y_{\mathrm{calc}} - y_{\mathrm{exp}}]^2}{\sum y_{\mathrm{exp}}^2} \right)^{1/2} \times 100 \quad (5)$$

where y_{calc} is EC_{50} found by PLSR, y_{exp} the activity value found by DPPH assay; summation extends to all the prediction samples [33]. The RSEP value obtained using enhanced fingerprints was 7.8%, while that calculated from reported data in reference [13], using chromatograms at 254 nm, was 20.5%.

The first latent variable explains 91.3% of Y (EC_{50}) variance, suggesting that LV1 contains the most information to predict antioxidant activity. Fig. 6a and b shows the loadings plot of LV1 and a plot of variable importance in projection (VIP) scores for EC_{50}, respectively. Those variables that have an important loading contribution and high VIP value belong to compounds that are highly related with antioxidant activity. However, if these variables increase or decrease the predicted EC_{50} values cannot be defined, because negative loading values do not necessarily mean that those concerned variables have a negative influence on the predicted result; so, it is not possible to assign an antagonist or synergist role to the peaks related to those variables. After reconstructing the chromatograms from their respective enhanced fingerprints, it was possible to assign the important variables found in the regression (VIP > 5) to specific retention times at each of the wavelengths. These retention times corresponded to peak apexes. Fig. 6c shows the four chromatograms of a sample; all the relevant peaks found studying Fig. 6a and b were marked on it, regardless of whether the peak is present or not in that particular sample. Recently, Perez-Meseguer et al. [2] purified from damiana extracts a compound that showed the best antioxidant activity in a TLC assay at 254 nm; it was identified as 8-C-β-[6-deoxy-2-O-(α-1-rhamnopyranosyl)-xylo-hexopyranos-3-uloside]. This is the only natural source from which this compound has been reported, and it was present in 41 native damiana samples analyzed and is one of those marked in this work as important at the same wavelength (RT of 19.40 min), but there are other compounds influencing the prediction of the EC_{50} value. Any compound marked as "important", at any wavelength, contributes significantly to the regression, so it deserves further investigation.

4. Conclusions

The enhanced chromatographic fingerprinting approach used in this work, combined with PLSR, produces very good results when predicting the antioxidant activity from complex chromatographic data of damiana samples. Contour plot is a useful tool for selection of proper working wavelengths in the construction of the enhanced fingerprints and suitable pretreatments may remove unwanted baseline variations. The alignment of complex chromatographic data compiled from several wavelengths was possible using correlation optimized warping. Thus, the introduction of noise and mathematical artefacts was avoided.

The constructed PLSR model explains a large amount of X and Y variance. A study of the importance of the regression variables in scores projection and loadings of LV1 (by far the most important for explaining Y values) makes it possible to find which chromatographic peaks are most important for the regression at the wavelengths chosen, which is a clue for finding which compounds are related to antioxidant activity.

Acknowledgments

The authors are grateful to CONACYT-México for grant 53877, PAICYT (UANL) SA1644-07 for financial support and the Spain's Ministry of Science and Technology for partially funding this research within the framework of Project CTQ2007-26528. Ricardo Lucio thanks to CONACYT for the financial support in the form of a fellowship (308657) and the Alßan Office: "Supported by the Programme Alßan, the European Union Programme of High Level Scholarships for Latin America, scholarship No. E07D401919MX".

References

[1] G.B. Mahady, J. Nutr. 131 (2001) 1120S.
[2] J. Perez-Meseguer, A. Garza-Juárez, R. Salazar-Aranda, M.L. Salazar-Cavazos, Y.C. de la Torre Rodriguez, V. Rivas-Galindo, N. Waksman de Torres, J. AOAC Int. 93 (2010) 1161.
[3] J.R. Lucio-Gutiérrez, J. Coello, S. Maspoch, Food Res. Int. 44 (2011) 557.
[4] M. Laasonen, T. Harmia-Pulkkinen, C.L. Simard, E. Michiels, M. Rasanen, H. Vuorela, Anal. Chem. 74 (2002) 2493.
[5] M. Dumarey, A.M. van Nederkassel, E. Deconinck, Y. Vander Heyden, J. Chromatogr. A 1192 (2008) 81.
[6] WHO, Guidelines on Good Manufacturing Practices (GMP) for Herbal Medicines, 2007, ISBN: 978 92 4 154716 1.
[7] EMEA, CPMP/QWP/2819/00, Rev. 1 (2006).
[8] EMEA, CPMP/QWP/2820/00, Rev. 1 (2006).
[9] FDA, CDER/FDA, Guidance for Industry, Botanical Drug Products, 2004.
[10] B. Dejaegher, G. Alaerts, N. Matthijs, Acta Chromatogr. 22 (2010) 237.
[11] X. Fan, Y. Cheng, Z. Ye, R. Lin, Z. Qian, Anal. Chim. Acta 555 (2006) 217.
[12] D.K.W. Mok, F. Chau, Chemom. Intell. Lab. Syst. 82 (2006) 210.
[13] A. Garza-Juárez, M. de la Luz Salazar-Cavazos, R. Salazar-Aranda, J. Perez-Meseguer, N. Waksman de Torres, Planta Med. 77 (2011) 958.
[14] Y. Jiang, B. David, P. Tu, Y. Barbin, Anal. Chim. Acta 657 (2010) 9.
[15] A.M. van Nederkassel, M. Daszykowski, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1096 (2005) 177.
[16] J.R. Lucio-Gutiérrez, J. Coello, S. Maspoch, Anal. Chim. Acta 710 (2012) 40.
[17] R. Ramirez-Duron, L. Ceniceros-Almaguer, R. Salazar-Aranda, L. Salazar-Cavazos Mde, N. Waksman de Torres, J. AOAC Int. 90 (2007) 920.
[18] A. Garza-Juárez, N. Waksman-De-Torres, R. Ramírez-Durón, M. Salazar Cavazos, Acta Chromatogr. 21 (2009) 217.
[19] L. Alcaraz-Meléndez, S. Real-Cosío, M.L. Robert, Sci. Hortic. 96 (2002) 293.
[20] L. Alcaraz-Meléndez, J. Delgado-Rodríguez, S. Real-Cosío, Fitoterapia 75 (2004) 696.
[21] R. Salazar, M.E. Pozos, P. Cordero, J. Perez, M.C. Salinas, N. Waksman, Pharm. Biol. 46 (2008) 166.
[22] S. Dudonné, X. Vitrac, P. Coutière, M. Woillez, J.M. Mérillon, J. Agric. Food Chem. 57 (2009) 1768.

[23] O. Blokhina, E. Virolainen, K.V. Fagerstedt, Ann. Bot. 91 (2003) 179.
[24] S.J. Leu, Y.P. Lin, R.D. Lin, C.L. Wen, K.T. Cheng, F.L. Hsu, M.H. Lee, Biol. Pharm. Bull. 29 (2006) 740.
[25] T. Skov, F. van den Berg, G. Tomasi, R. Bro, J. Chemom. 20 (2006) 484.
[26] G. Tomasi, F. van den Berg, C. Andersson, J. Chemom. 18 (2004) 231.
[27] R.G. Brereton, Pattern Recognition, Applied Chemometrics for Scientists, John Wiley & Sons, Ltd, 2007, 145 pp.
[28] H. Martens, T. Naes, Multivariate Calibration, Wiley & Sons, Great Britain, 1989.
[29] R.G. Brereton, Calibration, Applied Chemometrics for Scientists, John Wiley & Sons, Ltd, 2007, 193 pp.
[30] H. Mia, Robust Calibration, Practical Guide To Chemometrics, second edition, CRC Press, 2006, 167 pp.
[31] M. Daszykowski, B. Walczak, Trends Anal. Chem. 25 (2006) 1081.
[32] Eigenvector Documentation, using Cross-Validation, http://wiki.eigenvector.com/index.php?title=Using_Cross-Validation (accessed December 2011).
[33] M. Otto, W. Wegscheider, Anal. Chem. 57 (1985) 63.