ELSEVIER

Journal of Chromatography A, 909 (2001) 259–269

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Multivariate calibration methods for quantification in strongly overlapping capillary electrophoretic peaks

Sònia Sentellas, Javier Saurina*, Santiago Hernández-Cassou, M. Teresa Galceran, Lluís Puignou

Department of Analytical Chemistry, University of Barcelona, Diagonal 647, E-08028 Barcelona, Spain

Received 25 May 2000; received in revised form 26 September 2000; accepted 19 October 2000

Abstract

Capillary zone electrophoresis with diode-array detection was applied to the separation of ebrotidine and its metabolites. However, three of these, which are neutral in the conditions studied, co-migrated with the electroosmotic flow signal. Therefore, strongly overlapping peaks were observed. The main aim of this study was to show the potentiality of capillary electrophoresis in combination with chemometrics. Multivariate calibration methods were applied to quantify these analytes in synthetic mixtures. The results obtained using partial least squares (PLS) were in agreement with actual values, with an overall prediction error of 9.7%.

Keywords: Multivariate methods; Regression analysis; Chemometrics; Ebrotidines

1. Introduction

Selectivity of detection in analytical chemistry can be achieved by separation techniques, which physically separate the compounds present in the sample. Capillary zone electrophoresis (CZE) has proved to be an excellent technique for the separation of charged compounds. CZE takes advantage of a higher separation efficiency, faster analysis time and lower consumption of reagents and samples than high-performance liquid chromatography (HPLC).

Ebrotidine is a histamine H_2-receptor antagonist used as antisecretory drug in the treatment of gastric ulcer [1]. The analysis of ebrotidine and its metabolites in body fluids is fundamental to understand its pharmacokinetics, distribution process and metabolism. For this purpose, analytical methods based on HPLC with UV–Vis detection [2,3] and mass spectrometry [4,5] were proposed by Rozman et al.

Alternatively, the separation of ebrotidine and its metabolites by CZE has been studied by Sentellas et al. [6]. Although separation was successful for all charged compounds, the separation of three neutral metabolites remained unsolved as these analytes strongly co-migrated and, furthermore, overlapped with the electroosmotic flow (EOF) signal. These neutral compounds, 4-bromobenzenesulfonamide (compound A), *N*-(2-methylsulfonylethylaminemethylene)-4-bromobenzenesulfonamide (compound B) and *N*-(2-methylsulfinylethylaminemethylene)-4-bromobenzenesulfonamide (compound C), are especially relevant as they are commonly found in human urine [2,3] (see Fig. 1 for their structures).

*Corresponding author. Tel.: +34-93-403-4445; fax: +34-93-402-1233.

E-mail address: xavi@apolo.qui.ub.es (J. Saurina).

0021-9673/01/$ – see front matter
PII: S0021-9673(00)01060-8

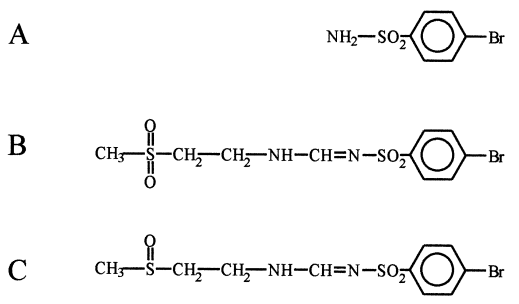

Fig. 1. Scheme of ebrotidine metabolites under study. Assignment: A = 4-bromobenzenesulfonamide; B = *N*-(2-methylsulfonylethylaminemethylene)-4-bromobenzenesulfonamide; C = *N*-(2-methylsulfinylethylaminemethylene)-4-bromobenzenesulfonamide.

The potential of CZE for the separation of uncharged compounds is obviously limited. However, when physical separation of sample components is not fully accomplished, quantification difficulties stemming from poorly-resolved peaks may still be overcome mathematically using chemometrics. The application of chemometrics to capillary electrophoresis has seldom been described. For instance, artificial neural networks have been used for optimization of the separation conditions [7–10] and chemometric techniques based on multivariate curve resolution using alternating least squares [11,12] and partial least squares (PLS) [13] have been satisfactorily used for quantification.

The paper is focused on the possibilities of multivariate calibration methods in the improvement of the quantification in poorly separated peaks from capillary electrophoretic runs. Of course, when developing a CE method, an effort should be made for optimizing the separation conditions in order to resolve all components present in the sample. However, the possibilities of the CE techniques are obviously limited and, sometimes, the separation is not completely successful. It is well known that CZE is not an appropriate technique for the analysis of poorly ionized compounds. Instead, other CE modes such as capillary electrochromatography and micellar electrokinetic capillary cromatography would be recommendable. This drawback might be partly circumvented using chemometrics for a mathematical separation of these compounds. Hence, the applicability of CZE could be expanded to the determination of poorly ionized compounds together with charged compounds in a single electrophoretic run. However, the chemometric procedures have also limits with respect to the number of compounds that can be analyzed simultaneously and the degree of overlapping of the spectral and electrophoretic responses. Therefore, multivariate calibration methods may be applied to quantify a few number of analytes with rather dissimilar electrophoretic or spectral profiles.

In this study, PLS is proposed for quantifying the neutral ebrotidine metabolites in several samples from the analysis of the corresponding CZE with diode-array detection (CZE–DAD) data. There is a great interest in the study of CZE data as they often show problems arising from the rather low reproducibility and accuracy in comparison with HPLC techniques, typically one order of magnitude higher. Then the analysis and quantification of compounds in CZE might be limited by these experimental variabilities. Either spectra or electrophoretic profiles (or even more complex signals) can be used as multivariate data arranged in data vectors for further mathematical analysis. Here, data pre-treatment to correct baseline and spectral drifts and peak shifting has been developed. The quantitative results confirm that PLS can overcome the lack of selectivity of the overlapping CZE responses of neutral ebrotidine metabolites.

2. Theory

Although multivariate calibration methods have been extensively described elsewhere [14–16], a brief description is given below. A common requirement of this type of multivariate calibration methods is that unknown samples and standards must have the same matrix composition in order to model all possible contributions to the response (i.e. analytes, interferences and matrix effects). For those samples with a complex matrix composition their standards should be, indeed, other samples analyzed beforehand by an independent method. Alternatively, synthetic standards can be used when the matrix is easily available.

2.1. Principal component regression (PCR)

PCR decomposes the experimental matrix of responses of the calibration set as follows:

$$\mathbf{R} = \mathbf{T}\,\mathbf{P}^{\mathbf{T}} + \mathbf{E} \quad (1)$$

R being the response matrix with a dimension $n_s \times n_w$ (number of standards by the number of working wavelengths); **T** the scores matrix ($n_s \times n_f$); $\mathbf{P}^{\mathbf{T}}$ the loading matrix ($n_f \times n_w$) (the superindex **T** indicates the transposed matrix); and **E** the residual error matrix ($n_s \times n_w$). n_f is the number of latent variables or factors included in the model which are able to keep the relevant variance of data.

Next, the scores matrix **T** is correlated with the concentration matrix **C** ($n_s \times n_a$), where n_a is the number of analytes, using the expression:

$$\mathbf{C} = \mathbf{TB} + \mathbf{E} \quad (2)$$

where **B** is the matrix of regression coefficients which is resolved by using a least squares procedure.

This model is subsequently applied to predict the concentrations of unknown samples.

2.2. Partial least squares regression

The PLS algorithm takes into account the information of responses and concentrations simultaneously [14,15]. There are two procedures available to solve the system: in PLS1, one model is built for each analyte by using its concentration vector (e.g. **C** is $n_s \times 1$), while in PLS2, all analyte concentrations are simultaneously considered in constructing the calibration model (e.g. **C** is $n_s \times n_a$). In both cases, factors from a PLS model are calculated as those variables that describe the maximum amount of relevant information from the spectral or electrophoretic response matrix and from the concentration matrix, as follows:

$$\mathbf{R} = \mathbf{T}\,\mathbf{P}^{\mathbf{T}} + \mathbf{E} = \Sigma \mathbf{t}_k \mathbf{p}_k + \mathbf{E} \quad (3)$$

$$\mathbf{C} = \mathbf{Q}\,\mathbf{S}^{\mathbf{T}} + \mathbf{F}' = \Sigma \mathbf{q}_k \mathbf{s}_k + \mathbf{F}' \quad (4)$$

where **R** is the response matrix; **T** and **P** are the score and loading matrices associated with the response (the superindex **T** indicates the transposed matrix); **C** the concentration matrix of the analyte; **Q** and **S** the scores and loading of the concentration matrix; and **E** and **F′** are the unexplained information from responses and concentrations, respectively. The corresponding scores and loading for the k^{th} factor are $\mathbf{t}_k$, $\mathbf{p}_k$, $\mathbf{q}_k$, and $\mathbf{s}_k$ ...

The inner relationship between responses and concentrations in the PLS model is given from their corresponding scores factor by factor:

$$\mathbf{q}_k = \mathbf{b}_k \mathbf{t}_k \quad (5)$$

where $\mathbf{b}_k$ represents the regression coefficients.

Scores and loading of each factor are not calculated simultaneously but consecutively. The residuals in the concentration matrix are minimized using the following equation, which is generalized for the k^{th} factor as:

$$\mathbf{F}_k = \mathbf{F}_{k-1} - \mathbf{t}_k \mathbf{b}_k \mathbf{s}_k^{\mathbf{T}} \quad (6)$$

where $\mathbf{F}_k$ and $\mathbf{F}_{k-1}$ are the residuals of concentrations not explained by a model with k and $k-1$ factors, respectively. Note that for the first factor $\mathbf{F}_0 = \mathbf{C}$.

2.3. Non-linear PLS (NL-PLS)

PCR and PLS algorithms have essentially been developed for modeling linear data since they apply inner linear relationships between responses and concentrations. However, the non-linearity of the response can be taken into account methods by including more factors. Alternatively to these linear methods, non-linear PLS procedures have been proposed, which utilize different types of inner non-linear relationships. As example, polynomial functions [16] have been used.

2.4. Prediction strategy

Once the model is built, it can be used to predict the concentration of unknown samples. A suitable strategy when working with small data sets is to predict each sample by using the remaining samples as standards to build the calibration model [17]. With this approach, the number of samples for the calibration was 25 with one for prediction.

The optimum number of latent variables needed to

build the calibration model was estimated by leave-one-out cross-validation [18], as the number that minimized the prediction error of the sum-of-squares (PRESS) function calculated as follows:

$$\text{PRESS}(k) = \sum_{i=1}^{\text{Samples}} (c_i - \hat{c}_i(k))^2 \quad (7)$$

where c_i is the actual concentration of analyte in the sample i and $\hat{c}_i(k)$ is the concentration predicted by multivariate calibration methods when considering k latent variables.

The accuracy of the model was evaluated by calculating the prediction error as follows:

$$\text{Prediction error (\%)} = \frac{\sqrt{\sum_{i=1}^{\text{Samples}} (c_i - \hat{c}_i)^2}}{\sqrt{\sum_{i=1}^{\text{Samples}} (c_i)^2}} \times 100 \quad (8)$$

3. Experimental

3.1. Chemicals

4-Bromobenzenesulfonamide (compound A), *N*-(2-methylsulfonylethylaminemethylene)-4-bromobenzensulfonamide (compound B) and, *N*-(2-methylsulfinylethylaminemethylene)-4-bromobenzensulfonamide (compound C) were purchased from Grupo Ferrer (Barcelona, Spain) and their structures are shown in Fig. 1. Stock solutions of each compound containing 2000 μg ml^{-1} were prepared in acetonitrile. Reagents used for the preparation of carrier electrolyte solution were acetic acid and ammonia (all from Merck, Darmstadt, Germany).

3.2. Samples

The one-, two- and three-component samples were prepared according to full factorial designs at two concentration levels. The uppercase letter indicates the metabolite and the number refers to the concentration level. The low level containing 100 μg ml^{-1} of analyte is referred to as 1; the high level containing 200 μg ml^{-1} is referred to as 2:

One-component samples: A1, A2, B1, B2, C1 and C2.

Two-component samples: A1B1, A1B2, A2B1, A2B2, A1C1, A1C2, A2C1, A2C2, B1C1, B1C2, B2C1 and B2C2.

Three-component samples: A1B1C1, A1B1C2, A1B2C1, A1B2C2, A2B1C1, A2B1C2, A2B2C1 and A2B2C2.

For instance, A1B2 is the two-component mixture composed of 100 μg ml^{-1} of compound A and 200 μg ml^{-1} of compound B.

3.3. Instrumentation

A Beckman P/ACE System 5500 (Fullerton, CA, USA) capillary electrophoresis system with diode array spectrophotometric detector was used. A 75 μm I.D. fused-silica capillary (Supelco, Bellefonte, PA, USA) of 57 cm total length (50 cm to the detector) was used. Data were processed with a PC using Beckman P/ACE station software (version 1.0).

3.4. Capillary electrophoresis conditions

The carrier electrolyte was 50 m*M* acetic-acetate buffer at pH 5.7 (adjusted with ammonia). The sample was hydrodynamically injected (0.5 p.s.i.; 1 p.s.i. = 6894.76 Pa) for 4 s. The separation voltage was 20 kV and the temperature was set at 25°C.

The capillary was activated by rinsing (20 p.s.i.) 1 *M* sodium hydroxide solution for 5 min, and then it was washed with ultrapure water for 5 min. The capillary was rinsed with the carrier electrolyte for 1 h at the beginning of the study and for 2 min before each run.

3.5. Data generation

Spectra were acquired in the range 190–290 nm at regular steps of 0.26 s during the electropherogram. Sixteen working wavelengths were chosen for analysis; every 5 nm in the range 220–270 nm, where the spectral information about the analytes was expected to be maximum, and every 10 nm for the remaining ranges. Data were converted into ASCII files for mathematical treatment.

3.6. Data pre-treatment

The data pre-processing was as follows:

3.6.1. Shift correction

The run-to-run variability of the migration time of the CZE peaks under study was corrected with a peak alignment procedure. The peak maximum at 220 nm was chosen for peak synchronization; at this wavelength, the response was mainly due to the solvent rather than to the analytes, and thus it was common to all samples without being affected by the analyte composition. In particular, a time window was used in which 40 time channels were taken after the peak maximum and 160 time channels before the peak maximum. All electrophoretic runs were synchronized using the expression:

$$t_{i\mathrm{syn}} = t_i - t_{\max} + 160 \quad (9)$$

where $t_{i\mathrm{syn}}$ is the synchronized time channel (for i ranging from 0 to 200); t_i is the original (un-synchronized) time channel and $t_{\max}$ is the time channel at which the peak maximum was reached. The resulting 201-channel window was between 6.50 and 7.37 min. Fig. 2b illustrate the synchronization of the original data shown in Fig. 2a.

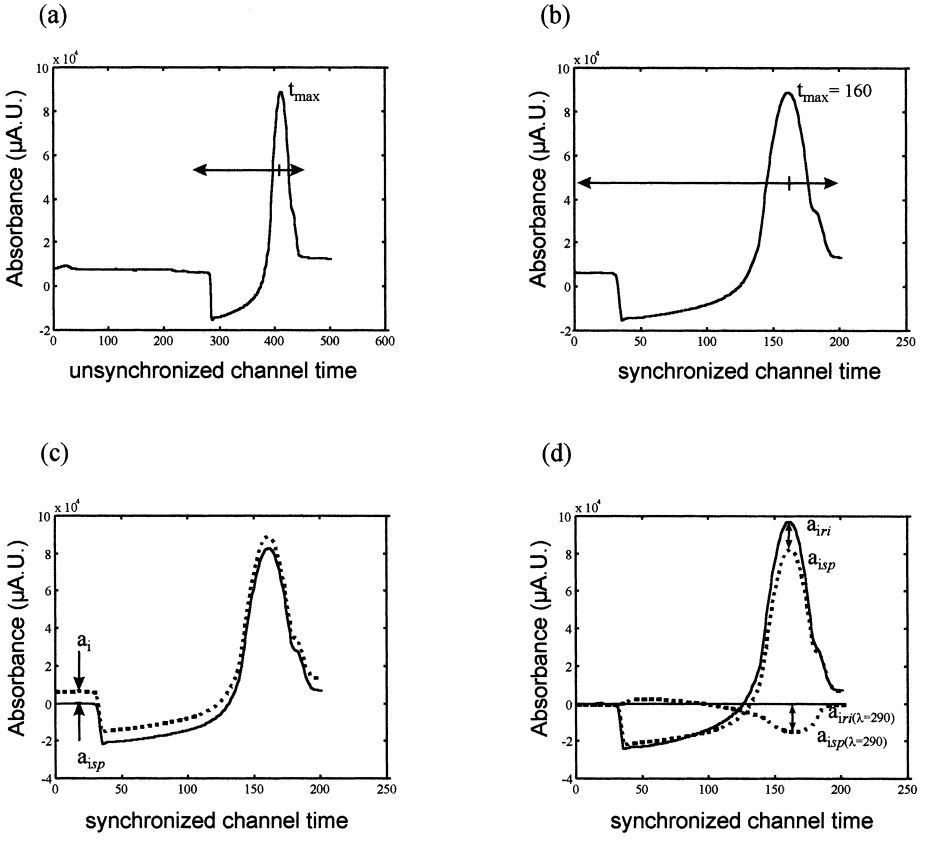

Fig. 2. Scheme of the data pre-treatment procedure. (a) Unsynchronized electrophoretic profile recorded at 220 nm. (b) Synchronized electrophoretic profile over a 201-time channel window; $t_{\max}$ was set at the 160 channel for all runs. (c) Spectral correction taking the absorbance at the beginning of the time window as a reference. (d) Refractive index correction using the absorbance at 290 nm as a reference. Legends: i refers to a given time channel; $a_{i\mathrm{sp}}$, spectral corrected absorbance; $a_{i\mathrm{ri}}$, refractive index corrected absorbance; $a_{i\mathrm{sp}(\lambda=290)}$, spectral corrected absorbance at 290 nm; $a_{i\mathrm{ri}(\lambda=290)}$, refractive index corrected absorbance at 290 nm. Dotted lines, input profiles (before being corrected); solid lines, output profiles (after being corrected).

3.6.2. Spectral correction

Taking the buffer background as a reference, the absorbance at the beginning of the time window should be zero at any wavelength. This is because, apart from the buffer, no absorbing compound is present in this zone. Possible spectral drift was corrected by subtracting the spectrum at time channel 0 from the spectral response over time, as shown in Fig. 2c. This correction was performed to all wavelengths as:

$$a_{i\mathrm{sp}} = a_i - a_0 \quad (10)$$

where $a_{i\mathrm{sp}}$ is the corrected absorbance value at the synchronized time channel i; a_i is the original absorbance value at the same time channel; and a_0 is the original absorbance value at time channel 0.

3.6.3. Drift correction and/or refractive index correction

At 290 nm neither ebrotidine metabolites nor the solvent absorb. Consequently, the absorbance over time at this wavelength should theoretically be zero. Hence, the behavior of the absorbance at this particular wavelength reflects possible baseline drifts and changes in the refractive index, owing to the different composition of the carrier electrolyte and the sample solvent. These effects were circumvented by subtracting the experimental absorbances at 290 nm from the corresponding spectrum for each measurement time (Fig. 2d):

$$a_{i\mathrm{ri}} = a_{i\mathrm{sp}} - a_{i\mathrm{sp}(\lambda=290\ \mathrm{nm})} \quad (11)$$

where $a_{i\mathrm{ri}}$ was the refractive index corrected absorbances for the time channel i.

3.7. Software

MATLAB for Windows (Version 4.1) [19] was used for calculations. Multivariate calibration methods were from the PLS_Toolbox [20].

4. Results and discussion

4.1. Data arrangement

Fig. 3a shows the three-dimensional data peak obtained for a three component sample. From all these data either spectral or time vectors can easily be obtained for further multivariate calibration analysis. For instance, the continuous dark line indicates a spectrum taken at a pre-selected migration time (Fig. 3b) and the dotted dark line indicates an electrophoretic profile taken at a given wavelength (Fig. 3c).

When analyzing several samples each one can produce spectral or time vectors. Then, the data matrices studied with PLS were constructed by superimposing vectors and keeping wavelengths (Fig. 4a) or migration times (Fig. 4b) in common, so that each row corresponded to one sample. Three single spectral data matrices (referred as **S** matrices) were selected for analysis: **S160**, **S155** and **S100**, which were taken at time channel 160, 155 and 100, respectively: their dimensions were 26×16. Four time data matrices (referred as **T** matrices) were also selected: **T240**, **T245**, **T255** and **T265**, taken at 240, 245, 255 and 265 nm, respectively: their dimensions were 26×201.

Progressing on the complexity of the data matrices, other arrangements can be analyzed (see Fig. 4c). Several spectral matrices taken at different migration times were simultaneously analyzed from the corresponding row-wise augmented matrix (e.g. **[S160, S155]** and **[S160, S155, S100]**). For example, **[S160, S155]** was the augmented data matrix built from S160 and S155 keeping each sample at the same row. A similar strategy was followed with time matrices (e.g. **[T240, T245, T255, T265]**). Furthermore, combinations of spectral and time matrices were also available (e.g. **[S155, T255]**).

4.2. Spectral and electrophoretic features

The spectra of the three ebrotidine metabolites studied are shown in Fig. 5a. The spectrum of compound A is easily distinguishable from the others with a maximum at 240 nm, and it presents poor absorption at λ higher than 260 nm. In contrast, compounds B and C have almost identical spectra, and their absorption at λ higher than 260 nm is still high.

At the experimental pH (pH 5.7) used in this separation, none of these three analytes are charged,

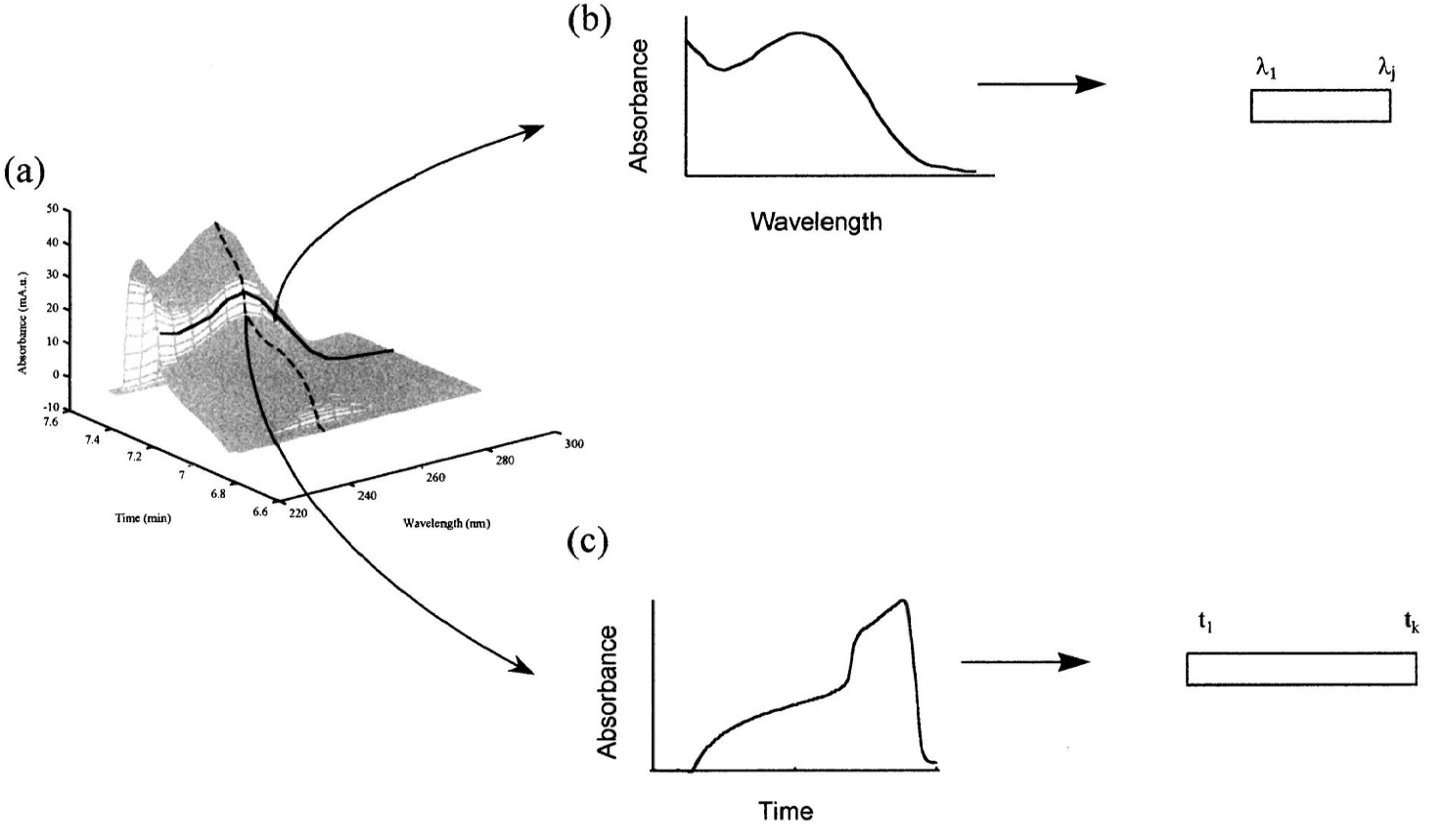

Fig. 3. Strategies for the generation of spectral and electrophoretic data vectors for PLS. (a) Three-dimensional plot of CZE–DAD run for the three-component mixture A1B1C1. (b) Spectrum at a preselected migration time. (c) electrophoretic profile at preselected wavelength.

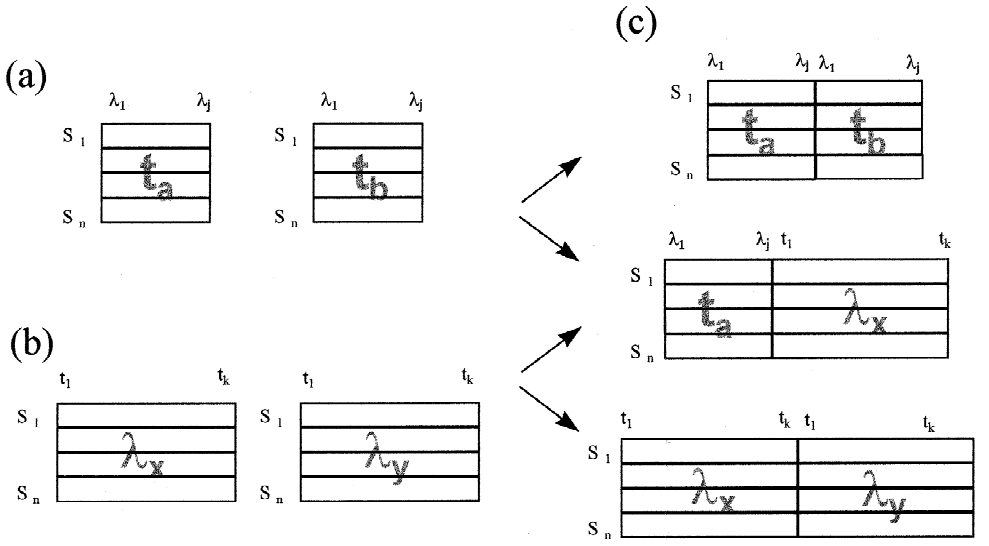

Fig. 4. Arrangements of spectral and electrophoretic data in single and row-wise (sample-wise) augmented data matrices. (a) single spectral matrices; (b) single-time matrices and (c) augmented data matrices.

so they co-migrated and overlapped with the EOF signal. Furthermore, as shown in Fig. 5b, the electrophoretic profiles of all the compounds were highly similar. The peculiar shapes of these peaks were attributed to the heterogeneous distribution of solvent and buffer inside the EOF zone. Acetonitrile was concentrated at the rear of the zone and then, as analytes have more affinity for acetonitrile than for the aqueous phase, they were also concentrated in this part.

No selective range was detected in any domain of measurement (spectral or time domain) and thus the quantification of such compounds from the CZE peak was difficult. Although this is a serious drawback when considering univariate calibration, this lack of selectivity can be overcome using multivariate calibration techniques.

4.3. Quantification with multivariate calibration methods

Preliminary studies were addressed to ascertain which spectral or electrophoretic data provided the best quantification strategy. For such a purpose, PLS2 was chosen as multivariate calibration method.

As indicate above, each sample concentration was predicted using the remaining 25 samples to build the calibration model. The estimation of the number of latent variables was based on minimizing the PRESS function, as illustrated in Fig. 6 for four examples of data matrices of different types. In this figure the arrows indicate the optimum number of latent variables. In these cases, among the contributions to data variance four factors could be attributed to the chemical components of the samples (compounds A, B and C and the EOF signal). Single spectral data matrices were successfully described with 5-factor models. Conversely, the number of factors to be included to model single time matrices and augmented data matrices was rather high, which indicated the great complexity of this type of data.

The predictive abilities of PLS2 models corresponding to different single and augmented spectral and time matrices are summarized in Table 1. Three single-spectral (**S100, S155, S160**) and four time (**T240, T245, T255, T265**) matrices were assayed for the PLS quantification. The best accuracy was obtained using spectral data as multivariate information and, in particular, the lowest prediction error was found for spectral data at time channel 155 (**S155**). These noticeable differences in the capabilities of each type of data for quantitative purposes were attributed to the higher dissimilarities in the spectral data which permitted the analytes to be distinguished

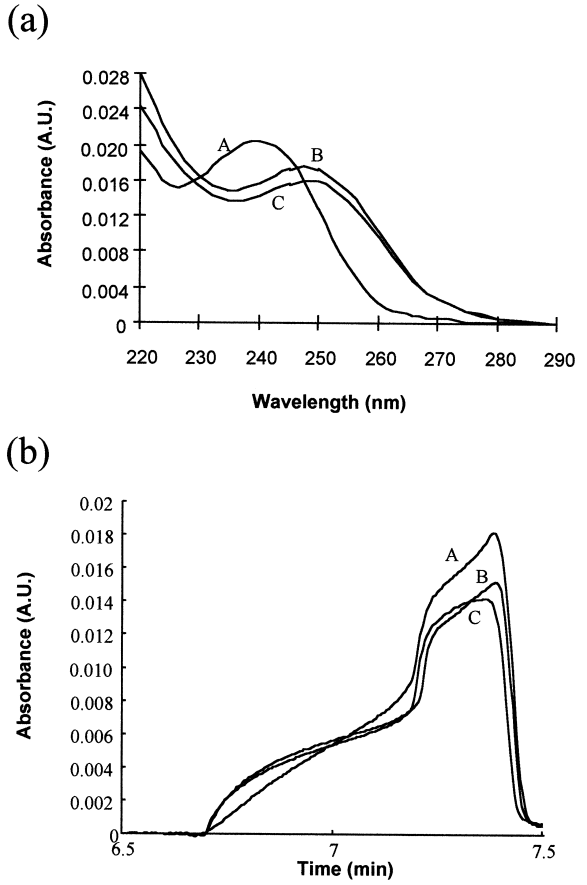

Fig. 5. (a) spectra and (b) electrophoretic profiles of compounds A, B and C from pure standards containing 200 μg ml^{-1}.

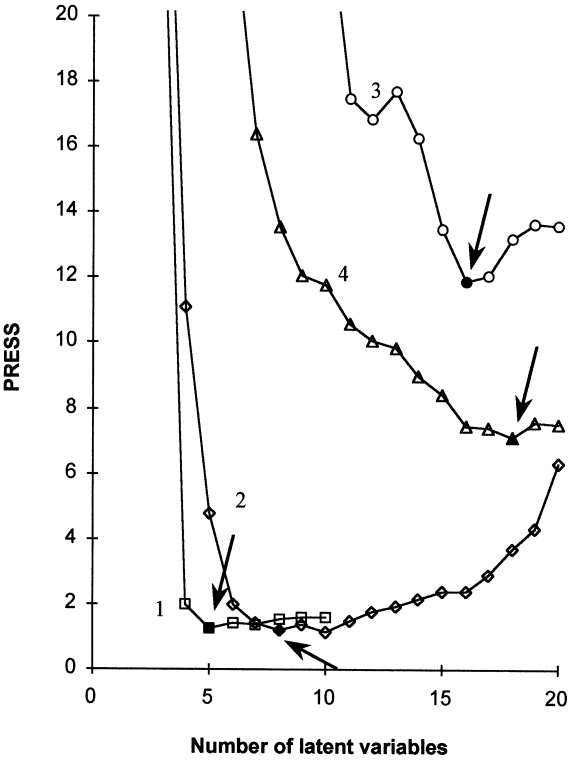

Fig. 6. Variation of PRESS values vs. the number of latent variables for different single and augmented data matrices. 1= **S155**, 2=**[S160, S155]**, 3=**T255** and 4=**[S155, T255]**.

Table 1
Percentage of prediction error in the quantification of compounds A, B and C in the set of samples from the analysis of several single and augmented data matrices using PLS

	Error (%)				
	Latent variables	Overall	A	B	C
S160	7	12.0	10.9	11.0	13.9
S155	5	9.7	7.6	10.7	10.3
S100	5	17.0	19.1	12.6	18.7
T240	7	46.8	43.3	53.5	42.9
T245	7	44.2	39.3	53.0	38.7
T255	16	29.7	27.2	29.9	31.9
T265	17	46.4	37.4	61.0	36.7
[S160, S155, S100]	12	11.0	9.2	10.4	13.1
[S160, S155]	8	9.3	7.8	8.7	11.0
[T240, T245, T255, T265]	20	28.9	27.5	37.2	19.0
[S155, T255]	18	23.0	27.7	21.8	18.6

more efficiently than using electrophoretic profiles. The variability of peak width was an additional source of variance that may lead to obtain poorer predictions.

Among the augmented data matrices, **[S160, S155, S100]**, **[T240, T245, T255, T265]** and **[S155, T255]** gave a poor predictions. The augmentation **[S160, S155]** was satisfactory but it did not significantly improve the quantification of the single matrix **S155**. In conclusion the analysis of the single-data matrix taken at the time channel 155 provided the most simple and appropriate data to carry out these analyses. These results suggested that the selectivity was not enhanced in these combinations since additional data matrices mostly provided redundant information.

The working wavelength range of spectral data was studied in order to find the optimum conditions for the quantification. Fig. 7 shows the variation of the prediction error against the wavelength. The best quantification was accomplished when using the intermediate working range from 220 to 290 nm. Prediction errors increased when widening or narrowing the range. This behavior can be explained as follows: on the one hand at the widest ranges the EOF marker contribution interfered with the response and made worse determinations, but it was almost negligible at wavelengths higher than 220 nm. On the other hand, at the narrowest ranges significant quantitative information concerning the analytes was lost and, thus, prediction errors increased.

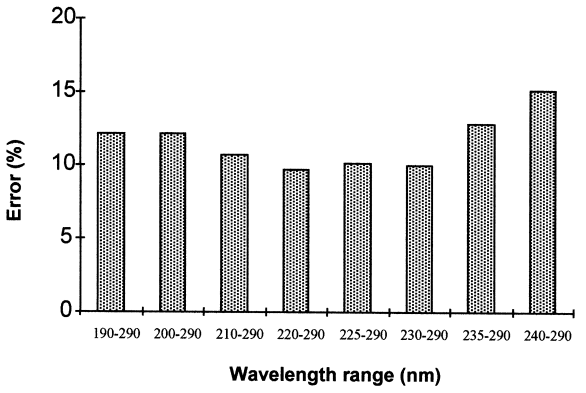

Fig. 7. Variation of the prediction error vs. the working wavelength range for the spectral data matrix **S155**.

Table 2
Percentage of prediction error in the quantification of compounds A, B and C using PCR, PLS1, PLS2 and NL-PLS[a]

	Error (%)			
	Overall	A	B	C
PCR	9.7	9.0	10.8	9.0
PLS1	9.5	8.9	10.7	8.9
PLS2	9.7	7.6	10.7	10.3
NL-PLS	11.2	10.7	11.0	12.0

[a] Conditions: single spectra data matrix taken at the time channel 155, **S155**.

Table 2 summarizes the results of the prediction of the sample mixtures with other multivariate calibration methods such as PCR, PLS1 and NL-PLS. The overall quantification errors were in all cases around 10%. The accuracy in the prediction was also similar for compound A, B and C. In the case of non-linear PLS modeling, the optimum degree of the

Table 3
Results of the quantification of compounds A, B and C in the set of samples with a 5-factor PLS model[a]

	C_A (μg ml^{-1})	C_B (μg ml^{-1})	C_C (μg ml^{-1})
a1	93	–	–
a2	197	–	–
b1	–	93	–
b2	–	194	–
c1	–	–	89
c2	–	–	197
a1b1	93	98	–
a1b2	90	194	–
a2b1	199	88	–
a2b2	192	189	–
a1c1	90	–	90
a1c2	95	–	196
a2c1	197	–	97
a2c2	191	–	186
b1c1	–	90	84
b1c2	–	90	192
b2c1	–	198	87
b2c2	–	169	174
a1b1c1	90	82	81
a1b1c2	80	100	182
a1b2c1	92	178	91
a2b1c1	185	89	77
a2b1c2	198	87	187
a2b2c1	194	191	82
a1b2c2	90	195	178
a2b2c2	179	194	184

[a] Conditions: single spectra data matrix taken at the time channel 155, **S155**. Working spectral range: 220–290 nm.

polynomial was two. When higher degrees were tested, data overfitting was observed and the prediction error increased considerably. This type of electrophoretic data was expected to behave linearly and, for this reason, non-linear algorithms did not improve the prediction given by linear algorithms.

Table 3 summarizes the results of the quantitative determination of ebrotidine metabolites A, B and C in the samples. These predictions were carried out under the modeling conditions selected, e.g. using the single spectral data matrix obtained at time channel 155 and for a working wavelength range from 220 to 290 nm. For each sample, a PLS model with five latent variables was built according to a leave-one-out cross validation strategy. Compound A was quantified more accurately than the others, with a mean prediction error around 7%; this was attributed to the higher dissimilarity of the spectrum of this compound. In the case of compound B and C, mean prediction errors were about 10–11%.

5. Conclusions

This paper has shown that multivariate calibration methods can improve the quantification in overlapping capillary electrophoretic peaks. The problem is illustrated with a particular application to quantify neutral compounds in complex samples. However, the general strategy can easily be extended to similar cases with overlapping peaks. Furthermore, this approach opens the possibility of determining in a single run a few number of neutral compounds together with charged analytes by CZE.

The chemometric analysis of the spectral or time data from CZE–DAD can be an attractive way for quantifying in poorly resolved peaks when the electrophoretic separation of analytes is not fully achieved. The best option for this determination seems to involve the use of spectral data rather than time data. This finding was attributed to the greater dissimilarities in the spectral profiles of the analytes under study. In this particular case, the performances of PCR, PLS1, PLS2 and NL-PLS were comparable.

Acknowledgements

This study was partially financed by DGICYT project PB96-0377 and AMB97-0405. Sònia Sentellas thanks the University of Barcelona for a grant (Beca de Formació en la Recerca i la Docència). The authors are obliged to Grupo Ferrer Internacional for providing ebrotidine and its metabolites.

References

[1] S.J. Kounturek, N. Kwiecien, E. Sito, W. Obtulowicz, K. Kaminski, J. Oleksy, Scand. J. Gastroenterol. 28 (1993) 1047.
[2] E. Rozman, M.T. Galceran, Ll. Anglada, C. Albet, J. Pharm. Sci. 83 (1994) 252.
[3] E. Rozman, M.T. Galceran, Ll. Anglada, C. Albet, Drug Metab. Dispos. 23 (1995) 976.
[4] E. Rozman, M.T. Galceran, C. Albet, J. Chromatogr. B 688 (1997) 107.
[5] E. Rozman, M.T. Galceran, C. Albet, Rapid Commun. Mass Spectrom. 9 (1995) 1492.
[6] S. Sentellas, L. Puignou, E. Moyano, M.T. Galceran, J. Chromatogr. A 888 (2000) 281.
[7] J. Havel, E.M. Peña, A. Rojas-Hernández, J.-P. Doucet, A. Panaye, J. Chromatogr. A 793 (1998) 317.
[8] M. Farkovà, E.M. Peña-Méndez, J. Havel, J. Chromatogr. A 848 (1999) 365.
[9] V. Dohnal, M. Farkovà, J. Havel, Chirality 11 (1999) 616.
[10] J. Havel, M. Breadmore, M. Macka, P.R. Haddad, J. Chromatogr. A 850 (1999) 345.
[11] R.M. Latorre, J. Saurina, S. Hernández-Cassou, Electrophoresis 21 (2000) 563.
[12] S. Sentellas, J. Saurina, S. Hernández-Cassou, M.T. Galceran, L. Puignou, Electrophoresis 22 (2001) in press.
[13] R.M. Latorre, J. Saurina, S. Hernández-Cassou, J. Chromatogr. A 871 (2000) 331.
[14] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, UK, 1989.
[15] K.B. Beebe, B.R. Kowalski, Anal. Chem. 59 (1987) 1007A.
[16] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Chemom. Intell. Lab. Sys. 7 (1989) 53.
[17] H.A. Martens, P. Dardenne, Chemom. Intell. Lab. Sys. 44 (1998) 99.
[18] S. Wold, Technometrics 20 (1978) 397.
[19] Matlab for Windows, Version 4.2, The Math Works, Natick, MA, USA, 1993.
[20] B. Wise, N.B. Gallagher, PLS_Toolbox 2.0 for use with Matlab, Eigenvector Research, Manson WA, USA, 1992.