Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Chromatography A, 1029 (2004) 13–20
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Peak purity determination with principal component analysis of high-performance liquid chromatography–diode array detection data

Kent Wiberg [a,b], Mattias Andersson [a], Anders Hagman [a], Sven P. Jacobsson [a,b,*]

[a] AstraZeneca R&D Södertälje, Analytical Development, Forskargatan 20, SE-151 85 Södertälje, Sweden
[b] Department of Analytical Chemistry, University of Stockholm, SE-106 91 Stockholm, Sweden

Received 16 October 2003; received in revised form 16 December 2003; accepted 18 December 2003

Abstract

A method is proposed for the determination of chromatographic peak purity by means of principal component analysis (PCA) of high-performance liquid chromatography with diode array detection (HPLC–DAD) data. The method is exemplified with analysis of binary mixtures of lidocaine and prilocaine with different levels of separation. Lidocaine and prilocaine have very similar spectra and the chromatograms used had substantial peak overlap. The samples analysed contained a constant amount of lidocaine and a minor amount of prilocaine (0.02–2 conc.%) and hence the focus was on determining the purity of the lidocaine peak in the presence of much smaller levels of prilocaine. The peak purity determination was made by examination of relative observation residuals, scores and loadings from the PCA decomposition of DAD data over a chromatographic peak. As a reference method, the functions for peak purity analysis in the chromatographic data system used (Chromeleon) were applied. The PCA method showed good results at the same level as the detection limit of baseline-separated prilocaine, outperforming the methods in Chromeleon by a factor of ten. There is a discussion of the interpretation of the result, with some comparisons with evolving factor analysis (EFA). The main advantage of the PCA method for determination of peak purity over methods like EFA lies in its simplicity, short time of calculation and ease of use.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Principal component analysis; Peak purity determination; Relative observation residuals; Scores; Loadings

1. Introduction

High-performance liquid chromatography with diode array detection (HPLC–DAD) is an important and widely used analytical technique of quantitative and qualitative analysis. In the pharmaceutical industry it is common to analyse the purity of drugs with HPLC, using DAD or mass spectrometric detection. The aim of purity analysis is to separate possible impurities from the main component and then, if possible, to identify them. These impurities can consist of degradation products, synthesis intermediates, packaging-related impurities, etc. In the analysis of the purity of a drug this often entails determining the purity of the main peak, which as a rule is the active substance, the reason being that if the main peak is not pure, it is possible for unresolved impurities to escape detection in the analysis, giving rise to incorrect analysis results. DAD with absorbance measured as a function of retention time and wavelength provides greater scope for analysing peak purity compared to single wavelength detection. Different methods of peak purity determination utilising DAD data have been developed, and modern chromatographic data systems generally have some functions for the analysis of peak purity in their software.

A number of chemometric methods for the analysis of overlapping chromatographic peaks have also been suggested. They include evolving factor analysis (EFA) [1,2] window factor analysis (WFA) [3], fixed-size window evolving factor analysis (FSW EFA) [4], heuristic evolving latent projections (HELP) [5,6] and eigenstructure tracking analysis (ETA) [7]. Many of these methods use principal component analysis (PCA) for decomposition of the data as a first step. One problem with applying PCA to matrices with a chemical rank larger than one is that it does not give directly interpretable profiles like chromatograms or spectra. The pure spectra cannot be found in a spectral

* Corresponding author. Tel.: +46-8-553-289-68; fax: +46-8-553-277-30.
E-mail address: sven.jacobsson@astrazeneca.com (S.P. Jacobsson).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2003.12.052

bilinear decomposition without external information. The explanation for this is that the score and loading vectors become linear combinations of the true analytical profiles since the scores and loadings are orthogonal in PCA. This issue is also called the problem of rotational freedom since the scores and loadings in PCA can be rotated without changing the fit of the model [8]. Hasegawa has previously shown that PCA can detect minute bands in an infrared reflection mixture spectrum [9]. PCA was shown to be capable of picking up small bands that were almost totally overlaid by the strong intensity bands in the raw spectra of a binary mixture in which one of the compounds was one hundred times smaller than the other. PCA did, therefore, provide a significant detection level for a chemical species with very low absorbance, even when the minute compound hidden in other dominant compounds could not be chemically separated [9].

The aim of the following study was to investigate whether a single PCA analysis of a sample analysed with HPLC–DAD could be used for determination of peak purity. The investigation of relative observation residuals, scores and loadings from a PCA decomposition of the LC-DAD data of a chromatographic peak should give information on the peak purity. In a pharmaceutical analysis of the purity of a drug the pure active compound is most often available. By comparing the results of a PCA decomposition of the DAD data for the main peak in the chromatogram under investigation with the corresponding PCA decomposition of a chromatogram of the pure compound, conclusions about peak purity can be drawn. This approach should also work with impurities in much lower concentrations than the active ingredient and with very low levels of chromatographic separation.

The method proposed was tested on synthetic binary mixtures of the local anaesthetic drugs lidocaine and prilocaine with different levels of separation. The focus was on determining the purity of the lidocaine peak and the samples analysed contained a constant amount of lidocaine and varying small amounts of prilocaine. The UV spectra of these two compounds are very similar, showing absorption at the same wavelengths, and the chromatograms used had substantial peak overlap. In the following text, the use of PCA for peak purity determination is denoted the PCA method. As a reference method, the method for peak purity analysis in the chromatographic data system used (Chromeleon) was applied.

2. Theory

The aim of chemometric methods for data decomposition and reduction is to find a small number of latent variables that can explain all the systematic information in the data matrix studied. PCA is a well-known chemometric method for the decomposition of two-way matrices [10]. The variance in the data matrix **X**, with m observations and n variables, is decomposed by successively estimating principal components (PCs) that capture the variance in the data in scores and loadings. Each calculated PC contains one score vector t, related to the observations, and one loading vector p', related to the variables:

$$\mathbf{X} = t_1 p_1' + t_2 p_2' + \cdots + t_R p_R' + E \tag{1}$$

After each calculated PC the unexplained variance is left in the residual matrix E. After the first PC has been calculated, the next one is calculated on the residual matrix E_1, which contains the variance not explained by the first PC. The variance of a PC is described by the eigenvalue, which is proportional to the variance explained. The eigenvalue can be described as the squared length of the PC and be estimated as the sum of the squares of the scores. From the residual matrix (E) the pooled residual standard deviation (s_0) for all observations in the data set can be calculated (valid for $m > n$):

$$s_0 = \sqrt{\frac{\sum_{k=1}^{m}\sum_{i=1}^{n} e_{ik}^2}{(n-r)(m-r-1)}} \tag{2}$$

where e_{ik}^2 is the residual of observation k at variable i, m the number of observations, n the number of variables and r is the number of PCs. The absolute residual standard deviation for a single observation is given by:

$$s_i = \sqrt{\frac{\sum_{i=1}^{n} e_{ik}^2}{n-r}} \tag{3}$$

The ratio s_i/s_0 shows the relative observation contribution and a well-explained observation will have a small value of the ratio, while the opposite will hold true for an unexplained observation. The ratio s_i/s_0 is called the relative observation residuals in the rest of this paper.

The data from a sample analysed with HPLC–DAD can be arranged in two ways if PCA is to be applied. The retention time can be set as observations (rows) and the wavelengths as variables (columns) or the opposite. In this study, the retention time was set as observations and the wavelength as variables since this is the approach chosen in EFA and related methods. If PCA is performed on a matrix of HPLC–DAD data arranged in this way, a plot of s_i/s_0 against retention time will reveal regions not as well explained as other regions by the current PC. Correspondingly, a plot of the scores versus retention time will show what part of the chromatogram the current PC has described and a plot of the loadings will show the spectral information explained. These three measures relative observation residuals, scores and loadings can be used to obtain information about peak purity.

In a mixture of two or more compounds at least two PCs will be needed to describe the data. The first PC explains to a greater extent the strongest absorbing compound, while the second PC will be more related to the second compound. Hence by calculating two PCs in a sample that is assumed to be pure, the presence of any other substance will show up in the second PC.

In modern chromatographic data systems different methods of assessing peak purity using DAD data have been employed. In this study, the functions for peak purity determination in the chromatographic data system Chromeleon were utilised as a reference method with which to compare the results of the PCA method. In Chromeleon the peak purity analysis can be done with a peak purity index (*PPI*) and peak purity match factor (*PPM*) [11]. The *PPI* is defined as the wavelength where the areas of the spectrum to the left and right are equal and thus represents the central wavelength of a spectrum. The *PPI* for a pure peak should then be the same from the start to the end of the peak for each spectrum recorded over it. By calculating the relative standard deviation (R.S.D.) of the *PPI* over a peak, it is possible to obtain a measure of the purity of the peak. The *PPM* describes the similarity of two spectral curves. It is calculated as the correlation between the spectrum at the peak maximum and the flanks of the peak. The method provides a measure of similarity ranging from identical (=1000) to different (=0). In an ideal case with a pure peak there would be no difference in the spectra at the start, the maximum or the end of the peak, and a 100% match factor would then correspond to a *PPM* of 1000 [11]. In this paper the *PPM* and R.S.D. of the *PPI* were used.

3. Experimental

3.1. Instrumentation

A Dionex Summit HPLC system with the chromatographic data system Chromeleon version 6.11 was used. The diode array detector used was a Dionex PDA 100 with a resolution of 1 nm and a data collection rate of 10 Hz. A Dionex ASI 100T autosampler and a Dionex P580 HPG high-pressure gradient pump were used. The column was μBondapak C_{18} and the mobile phase was phosphate buffer with a varying amount of acetonitrile. The flow was 1.0 ml min^{-1}. Samples were prepared by weighing into stock solutions with a calibrated balance, Sartorius MC5. The stock solutions were then further diluted with a Hamilton Microlab 1000, an automatic diluter for the different concentrations used in the study. PCA was performed with Simca-P version 9 (Umetrics, Sweden) and in this chemometric software the PCA decomposition is made with the NIPALS algorithm [12]. EFA calculations were made with Matlab version 6.0 (MathWorks, USA) with PLS toolbox 2.1 (Eigenvector Research, USA).

3.2. Reagents

Spectroscopic quality chemicals were used in this study. Mixing 1.3 ml of 1 M sodium dihydrogenphosphate with 32.5 ml of 0.5 M disodium hydrogenphosphate with water gave the phosphate buffer used. Acetonitrile was of analytical reagent grade. Lidocaine hydrochloride and prilocaine hydrochloride were supplied by AstraZeneca Bulk Production, Södertälje, Sweden. A Millipore Milli-Q filtration/purification system was used for production of the water used in the study. The UV spectra and structures of lidocaine and prilocaine are shown in Fig. 1.

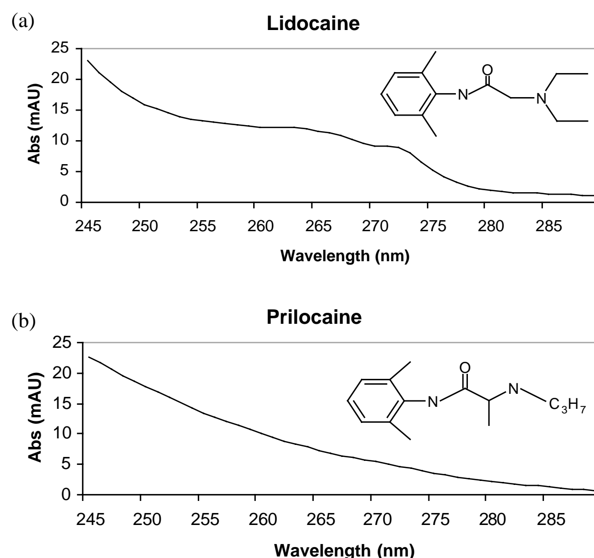

Fig. 1. Structure and UV spectra of lidocaine and prilocaine: (a) lidocaine and (b) prilocaine.

3.3. Analytical procedure

In this study three different levels of separation and resolution (R_s) were used: partly separated peaks Case 1 ($R_s = 0.7$) and Case 2 ($R_s = 1.0$) and baseline-separated peaks Case 3 ($R_s > 2$). Varying the amount of acetonitrile in the mobile phase (phosphate buffer) gave the different resolutions between the peaks. The retention time of the main lidocaine peak did hence vary a little in the three cases of separation.

For all three levels of separation, a test set of 10 samples containing binary mixtures of lidocaine and prilocaine as well as one pure lidocaine sample (0.667 mM) was analysed with the HPLC–DAD system. All ten binary mixtures contained 0.667 mM lidocaine and a minor amount of prilocaine from 0.02 to 2.0 conc.% (Table 1). Sample 1 did hence contain 0.667 mM lidocaine and 0.0133 mM prilocaine, $100(0.0133/0.667) = 2\%$. The results of these chromatographic analyses were then used for peak purity analysis, where the focus was on determining the purity of the large lidocaine peak in the presence of small amounts of prilocaine. For the peak purity analysis, the PCA method and the methods in Chromeleon were used. In all analyses the wavelength region used was 245–290 nm, with a resolution of 1 nm. This wavelength region was chosen since it is where lidocaine and prilocaine show absorption in the UV range (Fig. 1); it has also been used in previous studies [13–15]. The data collection rate used in the study was 10 Hz.

Table 1
Content of the samples in the test set

Sample no.	Amount prilocaine (conc.%)
Binary mixture of 0.667 mM lidocaine and varying amounts of prilocaine	
1	2.0
2	1.0
3	0.8
4	0.6
5	0.4
6	0.2
7	0.1
8	0.06
9	0.04
10	0.02

For the analysis of the partly separated peaks, the DAD data in the time windows 2.5–5 (Case 1, $R_s = 0.7$) and 3–5.5 (Case 2, $R_s = 1.0$) min were used giving data matrices with 1500 objects (rows) and 46 variables (columns).

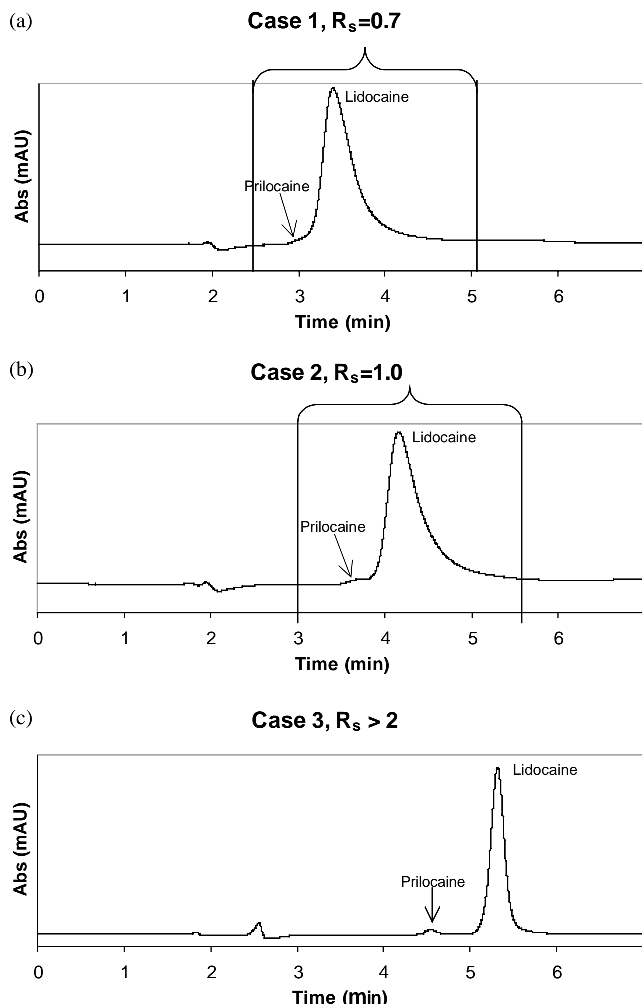

Fig. 2. Chromatograms and time window used in the data analysis for Cases 1–3 for sample 4 containing 0.6 conc.% prilocaine. Detection took place at 245 nm: (a) Case 1, $R_s = 0.7$; (b) Case 2, $R_s = 1.0$ and (c) Case 3, $R_s > 2$.

These time windows were chosen since this was where the peaks eluted. For the PCA method, one separate PCA model with two PCs was calculated on the mean-centred data of each sample, after which the relative observation residuals, scores and loadings were plotted and investigated. For the application of the peak purity methods in Chromeleon, all chromatograms were analysed in the same way. This meant integrating the peak in the time window mentioned above. The *PPM* and R.S.D. of *PPI* values so obtained for samples 1–10 with the two levels of separation were then examined.

For evaluation of the lowest detectable prilocaine amount, baseline-separated peaks of the samples were also analysed (Case 3, $R_s > 2$). This was done with the same chromatographic column but with a modified mobile phase. For these samples, no peak purities were determined, only the lowest detectable prilocaine peak in terms of peak height (three times the signal noise). The chromatograms and time window used in the data analysis for Cases 1–3 for sample 4 containing 0.667 mM lidocaine and 0.6 conc.% prilocaine are shown in Fig. 2.

4. Results and discussion

When PCA decomposition takes place of the DAD data of a chromatographic peak, the first PC will describe the strongest absorbing compound and the scores and loadings will show the chromatographic and spectral profiles of this compound. In the case of a pure lidocaine peak or when the content of lidocaine is much higher than other compounds, t_1 will thus resemble the chromatogram and p_1 the average UV spectrum of the peak (not shown here). The scores and loadings of the second PC (t_2 and p_2) will capture any other absorbing compounds in the peak. The scores of the first PC will capture the shape of the average chromatogram and the corresponding loadings the average spectrum. The scores and loadings of subsequent PCs will however show a more abstract shape since they are related to the variance left in the residual matrix of each PC. The results of the PCA method are shown in Figs. 3–5, where s_i/s_0 for the first PC as well as t_2 and p_2 are shown for samples 3, 5 and 7, partly separated peaks, Case 1 ($R_s = 0.7$). In these figures a comparison is made with the corresponding plots of s_i/s_0, t_2 and p_2 for a pure lidocaine peak. It should be elucidated that the variance explained in the second PC only is a fraction of that explained by the first PC. The first PC explained >99% of the variation in the data while the second PC explained <0.2%. This is the explanation for the small score values and s_i/s_0 in Figs. 3–5.

In Fig. 3, s_i/s_0 for the first PC and scores and loadings for the second PC (t_2 and p_2) are shown for sample 3 (solid lines), partly separated peaks, Case 1 ($R_s = 0.7$), and a comparison is made with the corresponding plots of the PCA decomposition of a pure lidocaine sample (dotted lines). Sample 3 contained 0.8 conc.% prilocaine and, as can be

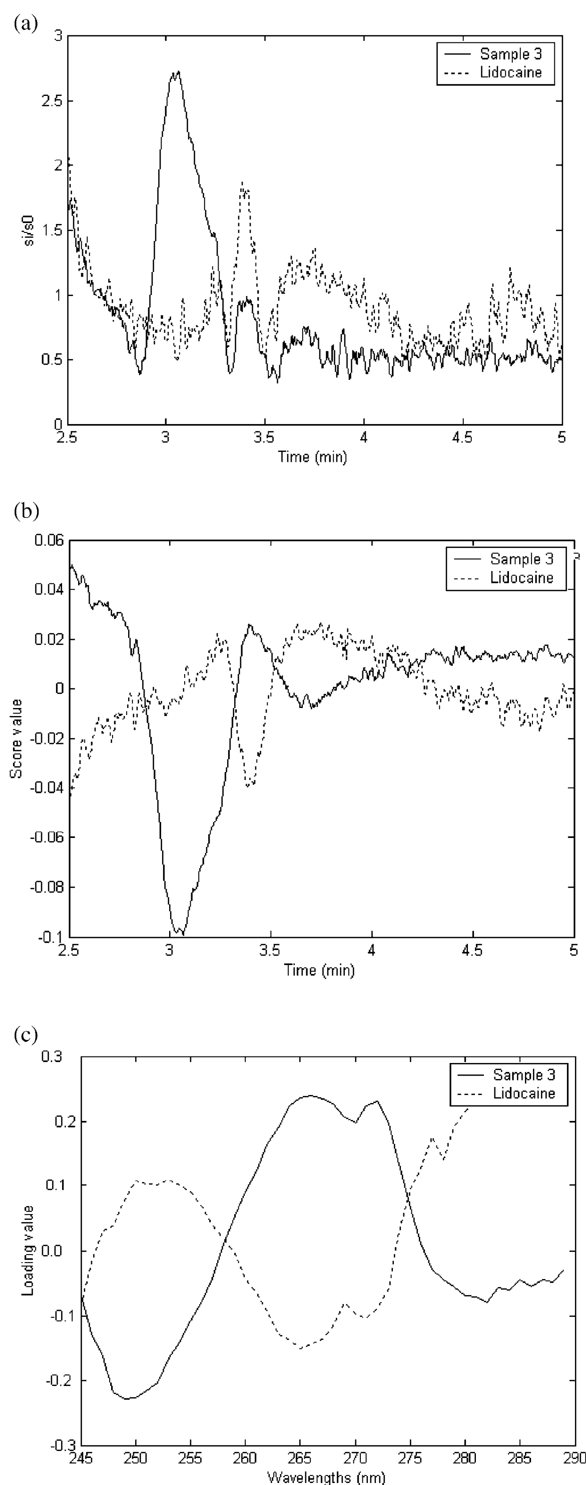

Fig. 3. Comparison of the result of PCA decomposition of sample 3 (0.8 conc.% prilocaine), partly separated peaks, Case 1 ($R_s = 0.7$), with the PCA decomposition of a pure lidocaine sample. The solid line refers to sample 3 and the dotted line the pure lidocaine sample: (a) s_i/s_0 after one calculated PC; (b) scores t_2 and (c) loadings p_2.

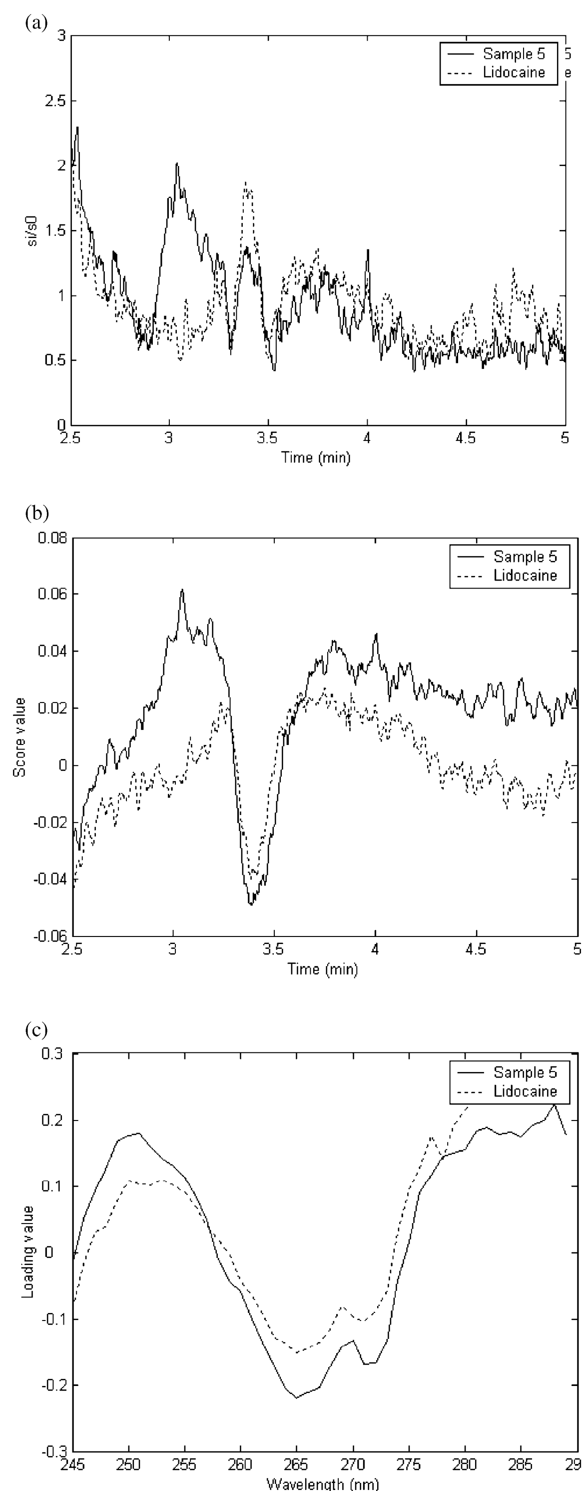

Fig. 4. Comparison of the result of PCA decomposition of sample 5 (0.4 conc.% prilocaine), partly separated peaks, Case 1 ($R_s = 0.7$), with the PCA decomposition of a pure lidocaine sample. The solid line refers to sample 5 and the dotted line the pure lidocaine sample: (a) s_i/s_0 after one calculated PC; (b) scores t_2 and (c) loadings p_2.

seen in Fig. 3a, s_i/s_0 has clearly captured a prilocaine peak at a retention time of about 3.1 min.

Looking at s_i/s_0 from the pure lidocaine sample, this peak is not present, although a small residual of lidocaine can be seen at about 3.4 min. In this pure lidocaine sample containing only one absorbing compound, the results in s_i/s_0 would have been expected to come only from noise since the first PC should have been enough to describe the vari-

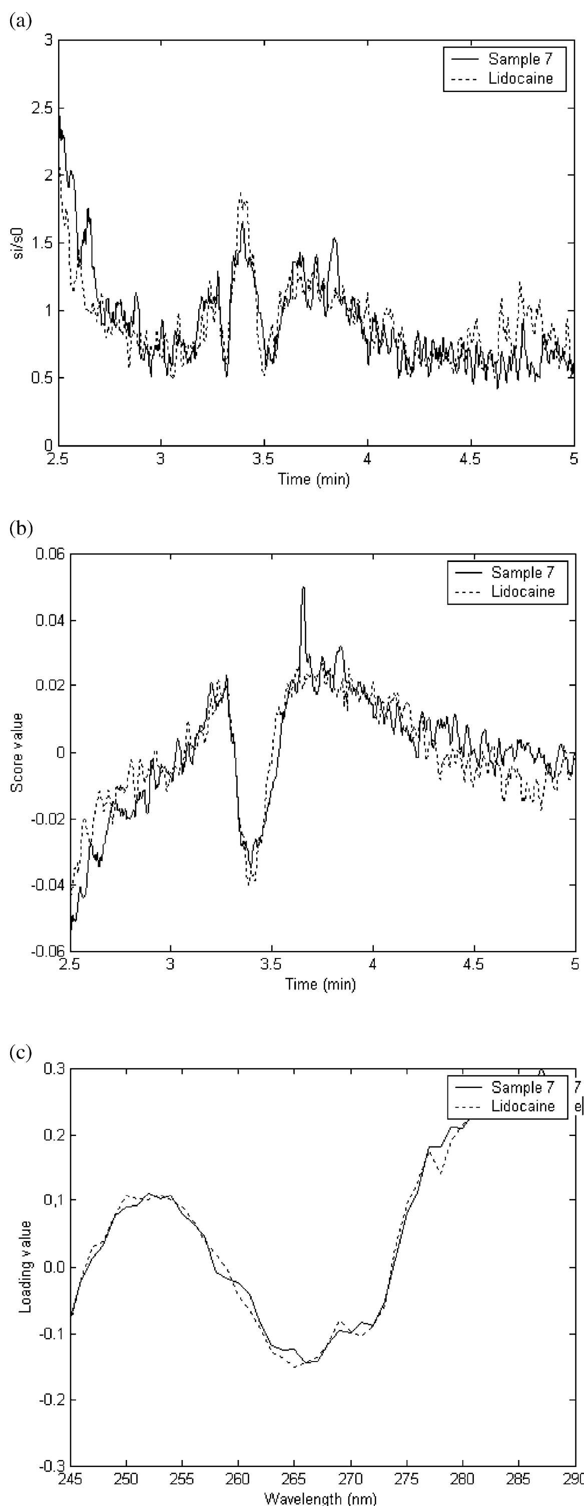

Fig. 5. Comparison of the result of PCA decomposition of sample 7 (0.1 conc.% prilocaine), partly separated peaks, Case 1 ($R_s = 0.7$), with the PCA decomposition of a pure lidocaine sample. The solid line refers to sample 7 and the dotted line the pure lidocaine sample: (a) s_i/s_0 after one calculated PC; (b) scores t_2 and (c) loadings p_2.

ance caused by lidocaine. However, as can be seen, this is not the case. Possible explanations for this might be a small level of non-linearity in the detector used or variations in the injection or chromatographic process. The magnitude of the residual lidocaine peak in s_i/s_0 was, however, exactly the same for pure lidocaine samples with different concentrations (not shown here).

In Fig. 3b, t_2 for the second PC is shown for sample 3 and the pure lidocaine sample. As can be seen, the results correspond with the s_i/s_0 in Fig. 3a, showing a prilocaine peak at about 3.1 min. In Fig. 3c the loadings (p_2) of sample 3 and a pure lidocaine sample are compared. As can be seen, the two lines differ and in fact look somewhat like mirror reflections of each other. The loadings p_2 of sample 3 resemble a lidocaine spectrum from which a prilocaine spectrum has been subtracted since p_2 captures the average spectrum left in the residual matrix E_1 after the first PC has been calculated. The shape of p_2 for the pure lidocaine sample is, however, a result of noise and residual lidocaine.

Fig. 3 thus clearly shows that the large lidocaine peak in sample 3 is not pure. In Fig. 4, s_i/s_0, t_2 and p_2 are shown for sample 5 (solid lines), partly separated peaks, Case 1 ($R_s = 0.7$), compared with the corresponding plots of the PCA decomposition of a pure lidocaine sample (dotted lines). Sample 5 contained 0.4 conc.% prilocaine. In Fig. 4a, s_i/s_0 for the first PC shows a small prilocaine peak at about 3.1 min. As can be seen in Fig. 4b, t_2 also shows some influence of a prilocaine peak at a retention time of about 3.1 min. However, the scores of the prilocaine peak at 3.1 min are now positive, in contrast to Fig. 3b. The explanation for this is that the second PC has rotated as a consequence of the very low prilocaine contribution in the scores t_2 (2.8–3.2 min). This rotation of the PC also causes the loadings p_2 to rotate (Fig. 4c), making them look like mirror reflections of the loadings for sample 3 (Fig. 3c). The influence of prilocaine is very small in sample 5 and noise and a residual of lidocaine cause the resulting shape of p_2.

The conclusion drawn from Fig. 4 is that the lidocaine peak is not pure (based on two replicate runs), although the influence of the small prilocaine amount is limited. The result of the PCA decomposition of sample 7, partly separated peaks, Case 1 ($R_s = 0.7$), containing 0.1 conc.% of prilocaine is shown in Fig. 5. Scores t_2, loadings p_2 and s_i/s_0 are shown by solid lines for sample 7 and by dotted lines for the pure lidocaine sample. The analysis of the ten samples with baseline separation showed that sample 6 containing 0.2 conc.% prilocaine was the sample containing the lowest amount of prilocaine that could be detected by the HPLC–DAD system used. As can be seen in Fig. 5a–c, there are no differences between s_i/s_0, t_2 or p_2 of sample 7 and the pure lidocaine sample. Hence it is concluded that the peak in sample 7 is pure according to the PCA method. The conclusion drawn from Figs. 3–5 is that for Case 1, $R_s = 0.7$, samples 3 and 5 were impure, but for sample 7, the peak was pure. The results obtained correspond well with the lowest detectable prilocaine peak obtained with baseline separation.

Table 2
Summation of the peak purity determination of the lidocaine peak

Method for peak purity determination	Level of separation	
	Case 1, $R_s = 0.7$ (conc.%)	Case 2, $R_s = 1.0$ (conc.%)
Chromeleon	2.0	2.0
PCA method	0.4	0.2

Results are given as the lowest conc.% of prilocaine in the binary samples analysed where the lidocaine peak could be determined as impure.

Table 2 shows a summation of the peak purity determinations of the lidocaine peak with the two different methods used at two different levels of chromatographic separation. With the methods in Chromeleon it was possible to determine the peak as impure containing 2.0 conc.% prilocaine in both Cases 1 and 2. With samples containing less prilocaine, the R.S.D. of PPI became <0.1 and the PPM 1000, indicating that the peak was pure according to the methods in Chromeleon. With the PCA method it was possible to correctly determine the purity of the lidocaine peak with 0.4 conc.% (Case 1) and 0.2 conc.% (Case 2) prilocaine. With the higher resolution the results were thus slightly better, as would have been expected. The results obtained for the PCA method thus outperformed the methods in Chromeleon and were at the same level as the detection limit of prilocaine with baseline separated peaks (sample 6, containing 0.2 conc.%).

4.1. Comparison with EFA

In EFA, PCA is applied to DAD data and the resulting eigenvalues are used for the analysis of unresolved chromatographic peaks [1,2]. EFA starts with PCA of the first two spectra and thereafter further PCA modelling takes place by adding one spectrum after another until eventually the whole matrix has been analysed (forward analysis). The procedure is then reversed, starting with PCA of the last two spectra and so on (backward analysis). As long as each spectrum added comes from the same compound, the data can be described with one PC, although if a second compound is detected, two PCs are needed to describe the mixture. The eigenvalues are plotted as a function of the retention time and the graphical results can in this way be used to determine the number of compounds in the mixture. In Fig. 6 EFA plots, log eigenvalues against retention time, are presented for samples 1, 3 and 5 (Case 1, $R_s = 0.7$) and the pure lidocaine sample. The EFA plots in Fig. 6 are presented with two PCs calculated in

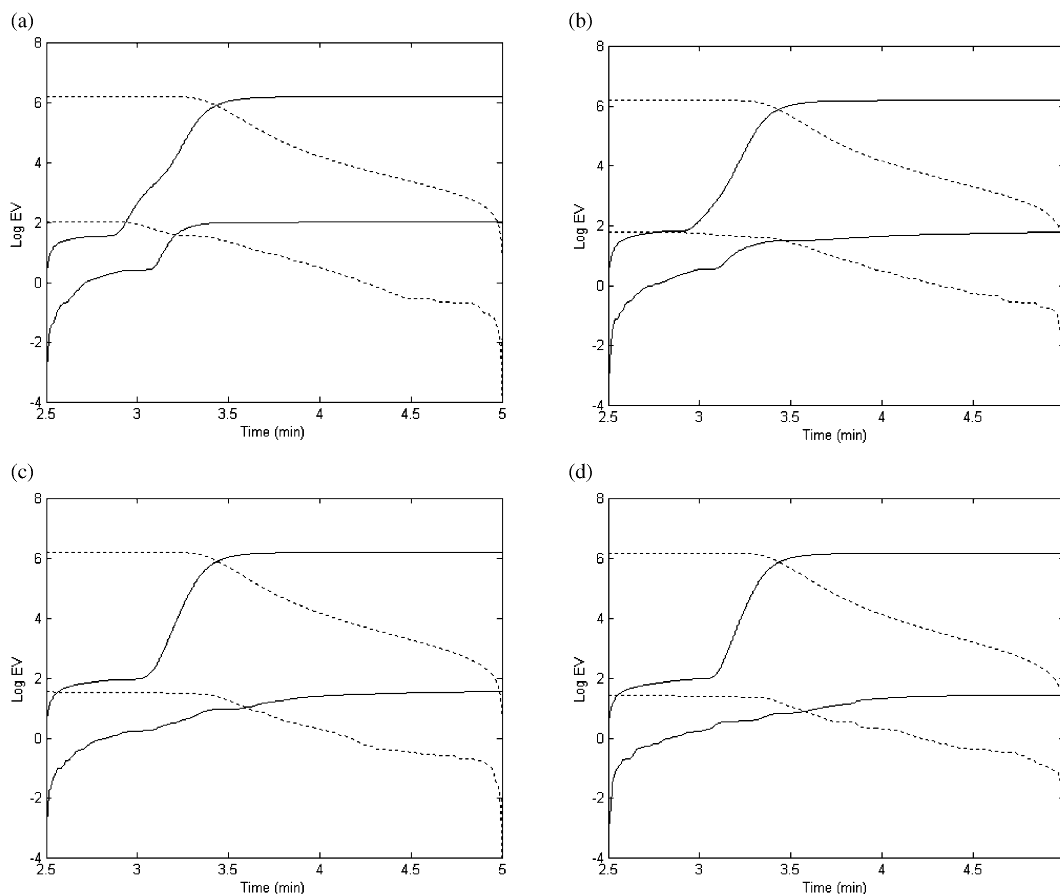

Fig. 6. EFA plot (log eigenvalue against retention time) for partly separated peaks, Case 1 ($R_s = 0.7$). Solid line: forward analysis and dotted line: backward analysis: (a) sample 3; (b) sample 5; (c) sample 7 and (d) pure lidocaine sample.

the same way as for the PCA method. The solid line represents the results of the forward analysis and the dotted line the backward analysis. New compounds appearing in the chromatogram are shown in this plot as eigenvalues rising from the background noise level. In Fig. 6a (sample 3) and b (sample 5) the elution of lidocaine and prilocaine is shown as where the solid lines rises from the background. The dotted lines show the end of the lidocaine peak dominating in these samples. For sample 7 (Fig. 6c), the prilocaine peak is no longer visible in the EFA plot. Comparing the EFA plots of samples 5 and 7 with the pure lidocaine sample, it can be seen that it becomes somewhat more difficult to distinguish a pure sample peak from the peak of a binary mixture where the concentration of one of the compounds is much lower than that of the other. The overall result of the EFA analysis does, however, correspond with the results obtained with the PCA method. The main advantages, however, of the PCA method proposed over methods like EFA for determination of peak purity lie in its simplicity, short time of calculation and ease of use. It is far easier and faster to perform only one single PCA on each sample under investigation than the multiple calculations entailed by methods like EFA. The PCA calculation on the data (1500 × 46) of a sample in this study with a standard personal computer (Pentium III, 700 MHz, 128MB internal memory) took <3 s while the EFA calculation on the same sample took >20 min. Performing the PCA method on a sample in this study was hence about 400 times faster than performing EFA and the peak purity information obtained with a single PCA was the same as with EFA. It would hence be possible to implement the PCA method as a method for real time monitoring of the purity of chromatographic peaks. By continuously calculating PCA on the DAD data of the eluting peaks in a chromatogram and plotting s_i/s_0 the peak purity could be determined.

From these results it can be concluded that by using the plots of relative observation residuals, scores and loadings from PCA performed on DAD data over a chromatographic peak, information regarding peak purity can be extracted. If a corresponding analysis of a sample containing only the main component can be made, this also enables a comparison with the results obtained for the unknown samples.

Possible factors influencing the proposed PCA method are the degree of spectral similarity of the partly separated compounds, the degree of chromatographic separation and instrumental parameters like signal to noise ratio of the diode array detector, etc.

For the PCA method, further PCs were calculated for all samples, although the conclusions drawn regarding the purity of the peaks remained the same as with two PCs. Calculating more than three PCs resulted in the modelling of noise. The PCA method was also tested without mean centring, although the results obtained were worse since the baseline noise then had a larger effect. One complication in this study was the fact that the UV spectrum of prilocaine is rather non-specific, showing no absorbance peak (Fig. 1b). The prilocaine spectrum has a leaning baseline kind of shape, not very different from the shape of noise. This caused the rotated loadings to have the same shape as the unrotated loadings even though the contribution from prilocaine was absent. However, the rotation of the PC could serve as an indication of the concentration level at which the second compound could no longer be detected. It is possible, moreover, that the scores (t_2) or relative observation residuals (s_i/s_0) could be used to estimate the concentration of the small prilocaine peak, although this was not tested in the present study. An alternative to the plot of s_i/s_0 is to plot the summed variable residuals ($\sum e_{ik}^2$) over time since this would result in the same shape as s_i/s_0. The PCA method was also tested on binary mixtures of lidocaine and 2,6-xylidine, with similar results.

5. Conclusion

The results of this study show that the PCA method, i.e. performing PCA of DAD data for a chromatographic peak with an investigation of relative observation residuals, scores and loadings, can be used to determine peak purity. The method proposed outperformed the reference method in Chromeleon. Furthermore, the proposed approach is far simpler and much faster to use than EFA, although the information obtained about whether or not a peak is pure is the same as in EFA and related methods. It should be stressed that the aim with the PCA method is not to determine the number of co-eluting compounds like EFA, rather the proposed method should be used to determine whether or not a chromatographic peak is pure.

References

[1] M. Maeder, A.D. Zuberbuehler, Anal. Chim. Acta 181 (1986) 287.
[2] M. Maeder, Anal. Chem. 59 (1987) 527.
[3] E.R. Malinowski, J. Chemom. 6 (1992) 29.
[4] H.R. Keller, D.L. Massart, Anal. Chim. Acta 246 (1991) 379.
[5] O.M. Kvalheim, Y. Liang, Anal. Chem. 64 (1992) 936.
[6] Y. Liang, O.M. Kvalheim, H.R. Keller, D. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946.
[7] J. Toft, O.M. Kvalheim, Chemom. Intell. Lab. Syst. 19 (1993) 65.
[8] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149.
[9] T. Hasegawa, Anal. Chem. 71 (1999) 3085.
[10] J.E. Jackson, A Users Guide to Principal Components, Wiley, New York, 1991.
[11] User's Guide to Chromeleon, Version 6.11, Dionex, Sunnyvale, CA, 2000.
[12] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1.
[13] K. Wiberg, A. Hagman, P. Burén, S.P. Jacobsson, Analyst 126 (2001) 1142.
[14] K. Wiberg, A. Hagman, S.P. Jacobsson, J. Pharm. Biomed. Anal. 30 (2003) 1575.
[15] K. Wiberg, M. Andersson, A. Hagman, S.P. Jacobsson, J. Pharm. Biomed. Anal. 33 (2003) 859.