ELSEVIER

Journal of Chromatography A, 912 (2001) 13–29

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

On the equations describing chromatographic peaks and the problem of the deconvolution of overlapped peaks

P. Nikitas*, A. Pappa-Louisi, A. Papageorgiou

Laboratory of Physical Chemistry, Department of Chemistry, Aristotle University of Thessaloniki, 54006 Thessaloniki, Greece

Received 28 July 2000; received in revised form 3 January 2001; accepted 3 January 2001

Abstract

The problem of the appropriate choice of the function that describes a chromatographic peak is examined in combination with the deconvolution of overlapped peaks by means of the non-linear least-squares method. It is shown that the majority of the functions proposed in the literature to describe chromatographic peaks are not suitable for this purpose. Only the polynomial modified Gaussian function can describe almost every peak but it is mathematically incorrect unless it is redefined properly. Two new functions are proposed and discussed. It is also shown that the deconvolution of an overlapping peak can be done with high accuracy using a non-linear least-squares procedure, like Microsoft Solver, but this target is attained only if we use as fitted parameters the position of the peak maximum and the peak area (or height) of every component in the unresolved chromatographic peak. In case we use as fitted parameters all the parameters that describe each single peak enclosed in the multi-component peak, then Solver leads to better fits, which though do not correspond to the best deconvolution of the peak. Finally, it is found that Solver gives much better results than those of modern methods, like the immune and genetic algorithms.

Keywords: Peak function; Peak resolution; Deconvolution methods

1. Introduction

Recently, three novel methods for deconvolving overlapped chromatographic peaks have been proposed [1,2]. All of them belong to the general case of computational methods that mimic natural processes. Thus, the immune algorithm of Ref. [1] imitates the defending process of an immune system, the genetic algorithm used also in Ref. [1] is based on Darwin's evolution rule and finally the artificial neural network used in Ref. [2] mimics the behaviour of the neural system in the brain.

The most important conclusion of these studies is that the development of new deconvolution methods for overlapping peaks is necessary due to serious limitations of the classical non-linear least-squares method [2]. However, this conclusion is not supported by other studies [3]. For this reason we examine this issue in the present paper in more detail, i.e., whether the non-linear least-squares method is inappropriate for deconvolution of overlapping peaks or there are conditions under which it can give correct results. Note that nowadays the non-linear least-squares method is so simple that it can be used not only by scientists but also by

*Corresponding author. Tel.: +30-31-997-773; fax: +30-31-997-709.

E-mail address: nikitas@chem.auth.gr (P. Nikitas).

0021-9673/01/$ – see front matter
PII: S0021-9673(01)00524-6

students. For example the simplicity and the capabilities of Solver, a programme for non-linear least-squares fitting using the spreadsheets of Microsoft Excel, have been pointed out in several articles [4–9]. On the contrary, the algorithms that mimic natural processes are still quite complicated and forbidden for the average analyst.

The deconvolution problem of overlapping peaks is closely associated with the appropriate choice of the function that describes the chromatographic peaks. In literature there is a good variety of such functions [1–3,10–17]. However, the majority of them are not general enough to describe every chromatographic peak. For example the exponentially modified Gaussian function (EMG) [10–13] is considered as one of the most effective functions for this purpose. Despite this we have observed that it cannot describe peaks characterised by tailing behaviour at the end of the peaks such as those obtained by electrochemical detection (ED) under special conditions. Thus, in the present paper we first examine the problem of the appropriate choice of the function that describes a chromatographic peak and then the deconvolution of overlapped peaks by means of the non-linear least-squares method.

2. Functions for chromatographic peaks

The most important functions used up to now to describe chromatographic peaks are the following:

1. Gaussian distribution function (GD) expressed as:

$$h(t) = h_m e^{-\left(\frac{t-t_m}{s}\right)^2} \quad (1)$$

where t is the time, $h(t)$ is the ordinate, i.e., the peak intensity, h_m is the height of the peak, t_m is the position of the peak maximum and s is a constant denoting the standard deviation of the Gaussian distribution.

2. Asymmetric Gaussian distribution function (AGD) [2]. It is expressed by Eq. (1) with $s = s_1$ when $t < t_m$ and $s = s_2 \neq s_1$ when $t > t_m$.

3. Generalised exponential function (GEX) [14]. If t_0 is the time where the detector output signal $h(t)$ starts to deviate positively from the baseline, then this function may be written as:

$$h(t) = h_m \left(\frac{t-t_0}{t_m-t_0}\right)^{b-1} \exp\left\{\frac{b-1}{a}\left[1 - \left(\frac{t-t_0}{t_m-t_0}\right)^a\right]\right\}, \; t > t_0 \quad (2a)$$

$$h(t) = 0, \; t \leq t_0 \quad (2b)$$

where a and b are constants.

4. Lorentzian function (L) [15]. It may be written as:

$$h(t) = \frac{h_m}{1 + (t-t_m)^2/s^2} \quad (3)$$

5. Polynomial modified Gaussian function (first case) (PMG1). It has been recently proposed by Torres-Lapasio and co-workers [3,17] and it is expressed by Eq. (1) with a polynomial standard deviation of the form:

$$s = s_0 + s_1(t-t_m) + s_2(t-t_m)^2 + \ldots \quad (4)$$

This function can describe every peak, it presents though the following defect. From a mathematical point of view $h(t)$ must tend to zero as t tends to infinity. However this property is violated by PMG1. When s is a second- or higher-order polynomial, $h(t \to \infty) = h_m$, whereas when s is a linear function of $t - t_m$, we have $h(t \to \infty) = h_m \exp(-1/s_1^2)$. This defect may cause problems in the deconvolution of overlapped peaks, because PMG1 may positively depart from the baseline outside the elusion region of each individual peak but inside the overlapping region of the unresolved peak. This is shown in the Results and discussion section. If we know this behaviour, we can easily overcome the defect by redefining the PMG1 function by means of Eqs. (1) and (4) within the elusion region and by $h(t) = 0$ outside this region. A different solution is proposed below.

6. Exponentially modified Gaussian function (EMG) [10–13]. This function arises from the convolution of the Gaussian function:

$$g(t) = \frac{A}{\sqrt{s2\pi\sigma}} e^{-(t-t_m)^2/2\sigma} \quad (5)$$

with the exponential function of unit area:

$$e(t) = e^{-\frac{t}{\tau}}/\tau \quad (6)$$

Therefore, we have:

$$h(t) = g(t) \otimes e(t) = \int_0^t g(x)e(t-x)\mathrm{d}x = \int_0^t g(t-x)e(x)\mathrm{d}x \quad (7)$$

which yields:

$$h(t) = Ae^q I/\tau \quad (8)$$

where:

$$q = \frac{\sigma^2}{2\tau^2} - \frac{t - t_\mathrm{m}}{\tau} \quad (9)$$

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\mathrm{d}x \quad (10)$$

and:

$$z = \frac{t - t_\mathrm{m}}{\sigma} - \frac{\sigma}{\tau} \quad (11)$$

As we have already pointed out the EMG function has been used widely to describe chromatographic peaks, nevertheless there are peaks that cannot be described by this function. The PMG1 function is more effective, however it exhibits the defect that $h(t \to \infty) \neq 0$. For these reasons we propose the following two functions, the capabilities of which will be examined in the next section.

1. Polynomial modified Gaussian function (second case) (PMG2). If we write the Gaussian function as:

$$h(t) = \frac{h_\mathrm{m} s_0}{s} e^{-\left(\frac{t - t_\mathrm{m}}{s}\right)^2} \quad (12)$$

where s is still given by Eq. (4), then $h(t \to \infty) = 0$ under all circumstances.

2. Generalised exponentially modified Gaussian function (GEMG). The chromatographic peaks that can not be described by the EMG function usually exhibit long tails indicating that the Gaussian function, Eq. (5), is not convoluted by a single exponential decay function but by the resultant of two exponential functions of different time constants. Such a combination with unit area may be expressed as:

$$e(t) = \frac{1}{1+b}\left\{e^{-t/\tau_1}/\tau_1 + be^{-t/\tau_2}/\tau_2\right\} \quad (13)$$

where b is the percentage contribution of the second exponential function to $e(t)$ function. If Eq. (13) is used instead of Eq. (6), then we readily obtain the generalised exponentially modified Gaussian function (GEMG):

$$h(t) = \frac{A}{1+b}\left\{e^{q_1} I_1/\tau_1 + e^{q_2} I_2/\tau_2\right\} \quad (14)$$

where:

$$q_i = \frac{\sigma^2}{2r_i^2} - \frac{t - t_\mathrm{m}}{\tau_i} \quad (15)$$

$$I_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{-x^2/2}\mathrm{d}x \quad (16)$$

and:

$$z_i = \frac{t - t_\mathrm{m}}{\sigma} - \frac{\sigma}{\tau_i} \quad (17)$$

3. Experimental

3.1. Chromatographic system and conditions

The liquid chromatography system consisted of a Shimadzu LC-9A pump, a Model 7125 syringe loading sample injector fitted with a 50 μl loop (Rheodyne, Cotati, CA, USA), a 250×4 mm MZ-Analytical column (5 μm Inertsil ODS-3), a Shimadzu UV–visible spectrophotometric detector (Model SPD-10A) and a Gilson ED system (Model 141) equipped with a glassy carbon electrode. The detector cell volume for UV was 8 μl and for ED 7.2 μl. The UV and ED systems were connected in series so that the analytes separated on the high-performance liquid chromatography (HPLC) column flowed first through the UV detector and then through the ED system. This allows dual measurement of analytes by UV absorbance and oxidation response at the ED system. Note that the ED system has been used in series with the UV detector because we observed that the chromatograms recorded by ED are much more asymmetric than those obtained by the UV detector at least when a mobile phase with a neutral

pH was used. Thus we have the chance to test the validity of the deconvolution method we propose here under extreme conditions.

The detection of the analytes was performed at 254 nm and 0.8 V vs. the Ag/AgCl reference electrode, respectively. The two detectors were interfaced to a Pentium personal computer (at 200 MHz) via a 14-bit AD-DA card. Laboratory-made software was used to record data and the measurements rate was 100 data points per min. The same computer was also used to carry out all calculations reported in this paper with programs written by the authors and/or with Microsoft Excel.

Since the mobile phase composition was systematically varied in this work (see below), the column was equilibrated with the new eluent for at least 15 min in order the baseline of the system to be stabilised. The volume flow-rate of all mobile phases used was 1.0 ml/min. All separations were carried out isocratically at ambient temperature.

3.2. Chemicals, standard solutions and mobile phases

All chemicals were used as received from commercial sources. 3-Methoxytyramine (3mt) hydrochloride, 5-hydroxytryptamine (5ht or serotonin) creatine sulfate complex, 5-hydroxytryptophan (5htp), *N*ω-methylserotonin (m5ht) oxalate salt, 5-hydroxyindole-3-acetic acid (5hiaa) and 4-hydroxy-3-methoxyphenyl-glycol (mhpg) hemipiperazinium salt were available from Sigma (St. Louis, MO, USA). All other chemicals were of analytical-reagent grade.

Catechol-related solutes were dissolved to a concentration of 100 μg/ml in water except mhpg which was initially dissolved to a concentration of 88 μg/ml. Working solutions of standards (1–20 μg/ml) were made by an appropriate dilution of the stock solutions. Solutes were injected individually or together using appropriate mixtures. The solute retention times and peak areas were obtained from the average value of at least three runs at each pH and percentage of organic modifier we used in this investigation. All solutions were kept refrigerated at 4°C when not in use.

Different mobile phases were prepared by combining an aqueous phosphate buffer (with constant ionic strength, 0.02 *M*, and pH value 3 or 7) with isopropanol to obtain solutions containing 1, 3 or 5% (v/v) organic modifier. All the eluents were filtered through a mixed ester membrane filter (0.45 μm, Schleicher & Schuell, Germany), sonicated and degassed under vacuum for 5 min before use.

3.3. Artificial data

Apart from the above experimental data, we also used synthetic data obtained by adding noise to the EMG equation, i.e., by means of the equation $h(t) = h_{EMG}(t) + \text{noise}$, where $h_{EMG}(t)$ is given by Eq. (8). The noise added to the EMG function was either a random noise of certain intensity or Gaussian noise with zero mean and a certain standard deviation.

3.4. Mathematical and computational treatment

The first treatment of the experimental data was the correction of the baseline by subtracting the linear interpolation in regions before and after the elution of the peaks. This treatment as well as the whole analysis of the experimental and the synthetic data has been carried out at Microsoft Excel spreadsheets. The Microsoft Solver, a powerful routine based on the Levenberg–Marquardt algorithm [18–20], was used for the non-linear least-squares fitting.

As has been pointed out previously, a fitting procedure by means of Solver is both simple and effective. For example, if we want to fit the EMG function to a chromatographic peak, we work as follows: We write in a column of four cells, say in cells B1:B4, first estimates of the values of A, t_m, σ and τ. Then at the seven first cells of any row below the fifth row we write the labels t, h(exp), q, z, I, h(calc) and SR. Below the cells with the labels t and h(exp) we introduce the experimental data and in the next column we calculate the quantity q by means of Eq. (9) using the values of parameters t_m, σ and τ which are at the cells B2:B4. For example, if the experimental data start from the sixth row, then at the cell C6 we write the function =($B$3/$B$4)^2/2−(A6−$B$2)/$B$4 and complete the column of q using the automatic filling procedure. Parameter z is similarly calculated, the integral I is calculated directly using the function NORMSDIST(z) and the EMG peak, h(calc), is calculated from Eq. (8). In the

last column the square of residuals (SR), i.e., the quantity $[h(\text{exp})-h(\text{calc})]^2$, is computed. Finally, the sum of squares of residuals (SSR) is calculated at some cell of the spreadsheet. At this stage it is quite useful to construct the plot of h(exp) and h(calc) vs. t. From the plot we can easily find the value t_m and we may also obtain a good initial estimate for the area A. For σ and τ we may start with the value 0.1. After that Solver is run to minimise the content of the cell with the sum of squares of residuals, SSR, by changing the contents of cells B1:B4, i.e., the cells that contain the initial estimates of fitted parameters A, t_m, σ and τ.

Solver was configured to use maximum time = 1000 s, iterations = 1000, precision = 10^{-12}, tolerance = 5%, convergence = 10^{-12}, automatic scaling, tangent estimates, central derivatives and Newton search.

We find more effective to run Solver in two steps. In the first step A and t_m are kept constant and the optimum values of σ and τ are obtained. After that Solver is run to find the best values of all parameters, A, t_m, σ and τ. Similarly, if the PMG1 function is used, then in the first step h_m, t_m are kept constant and Solver finds the optimum values of s_i and in the next step all parameters are treated as fitted parameters. Note that the Monte-Carlo technique we proposed previously [9] for finding best initial estimates for Solver is not necessary in the present case, unless we use many terms in the expression of s in the PMG1 and PMG2 functions.

The above procedure was followed for the analysis of an individual peak. A similar procedure was adopted for the deconvolution of overlapped peaks. The fitted function was the sum of n individual functions, where n is the number of the overlapped peaks. It is seen that the number of fitted parameters is now great enough. For example, if the GEMG function is adopted for the deconvolution of a three-component peak, the number of the fitted parameters is equal to $3\times6=18$. However we should point out the following: In chromatography and during the identification procedure the retention of the individual solutes of an overlapping or not chromatogram is always known by means of injection of standard solutions. Therefore, if we fit the GEMG or any other function to the peaks of the standard solutions, we obtain good initial estimates for all the fitted parameters except the area A or the height h_m, because these two parameters depend upon the solute concentration which may be different in the specimen than in the standard solution.

The final application of Solver or any other nonlinear least-squares procedure may follow two different approaches: in the first approach we assume that all parameters that determine the shape of an individual peak, i.e., the parameters σ and τ of the EMG function or the parameters σ, b, τ_1 and τ_2 of the GEMG function or the s_i parameters of the PMG1 and PMG2 functions, have the same values in the multi-component peak as in the single peak. That is, we may assume that the profile of the chromatographic peaks of the standard solutions does not change in the unresolved peak [21]. If we adopt this assumption, then the fitted parameters are limited to the retention times (t_m) of the individual solutes and either to the areas (A) or heights (h_m) of the individual peaks depending on the choice of the fitted function. Note that it was proven in this investigation that small retention times shifts need to be considered when a deconvolution technique is applied. That is, the retention time of an individual solute may be slightly different in the multi-component peak than in its single peak.

In the second approach we do not accept the assumption of the first approach about peak parameters and the number of the fitted parameters is n times the number of the parameters of each individual peak. For simplicity the first approach of fitting some parameters will be denoted by FSP, while the second approach of fitting all peak parameters will be denoted by FAP. These two approaches do not give in general similar results, as we shall see in the next section.

4. Results and discussion

4.1. Choice of the appropriate function

The sum of squares of residuals (SSR) after the fitting can be used as a criterion for the choice of the function that describes satisfactorily the experimental chromatographic peaks. The values of SSR for the various solutes used in the presence of 5% isopropanol in neutral and acid mobile phases are given

Table 1
Values of SSR obtained by means of the EMG, GEMG, PMG1, PMG2, AGD, GEX and L functions at mobile phases consisting of 5% isopropanol

pH	Solute	EMG	GEMG	PMG1	PMG2	AGD	GEX	L
ED system								
3	5ht	2054.7	106.5	172.2	133.1	7263.1	4895.8	15 223.4
3	5htp	1671.5	184.3	303.9	296.2	5890.0	4418.5	19 304.8
3	m5ht	868.6	40.6	71.9	64.1	3061.5	2149.8	7717.0
UV detector								
3	5ht	0.085	0.082	0.030	0.030	0.357	0.178	1.683
3	5htp	0.079	0.079	0.057	0.057	0.375	0.251	8.692
3	m5ht	0.104	0.104	0.096	0.096	0.240	0.171	1.613
ED system								
7	5ht	1943.9	252.8	8.92	9.9	13 452	14 448	2095.3
7	5htp	846.5	3.1	65.4	47.1	4887.7	2973.6	201.9
7	3mt	1519.1	97.4	2.37	4.2	9149.8	5539.4	485.7
7	mhpg	33.4	2.1	11.3	11.6	351.6	2386.7	1025.4
7	5hiaa	71.6	16.8	13.6	14.4	381.3	224.3	1012.0
UV detector								
7	5ht	0.076	0.041	0.024	0.023	1.245	1.908	0.710
7	5htp	0.083	0.014	0.043	0.045	0.748	11.60	0.363
7	3mt	0.041	0.007	0.009	0.007	0.249	0.247	0.069
7	mhpg	0.034	0.034	0.033	0.033	0.039	0.038	0.085
7	5hiaa	0.134	0.134	0.047	0.047	0.290	0.221	14.42

in Table 1. Table 2 includes the values of SSR in neutral mobile phases in the presence of 1% and 2% isopropanol.

In respect to these tables we should clarify the following: (a) the values of SSR are obtained when the peak intensity is measured in nA for the ED peaks and in mAU for the UV peaks. (b) Parameter *s* in the Lorentzian function was a polynomial of sixth-order of the form of Eq. (4). The parameter s in the Lorentian function is usually considered as constant. Here, we adopted the polynomial expression of Eq. (4) to increase the applicability of this function. (c) The PMG1 and PMG2 functions were fitted to the UV peaks using a fourth-order polynomial for *s* and this polynomial was increased to a sixth order to fit the ED peaks, because the ED peaks are usually highly asymmetric.

From Table 1 we observe that the AGD, GEX and L functions fail to describe our chromatograms. For this reason they have not been used further in this work. Thus, in Table 2 the fitted functions are the EMG, GEMG, PMG1 and PMG2. It is also seen that the EMG function describes fairly well all the UV peaks but it fails for the peaks of the ED system, except for the peaks of mhpg and 5hiaa. The GEMG, PMG1 and PMG2 functions describe all chromatographic peaks accurately. However, this accuracy for the PMG1 and PMG2 functions is usually achieved using more fitted parameters than the GEMG function. Moreover, as we have already pointed out, the behaviour of the PMG1 function over the entire overlapping region may be erroneous. This is clearly depicted in Fig. 1, where the PMG1 and PMG2 functions have been fitted to the chromatographic peak of mhpg obtained by ED in the presence of 1% isopropanol using a sixth-order polynomial for *s*. Note that the overlapping region in this case is from $t=9$ min to $t=18$ min (see Fig. 3A below). We observe that the PMG1 function increases abruptly just outside the elusion region of the individual peak and tends to get the limiting value h_m. The PMG2 function does not exhibit so pronounced deviations from the baseline outside the elusion region and in fact $h(t)$ tends to zero as t tends to infinity. However, small irregularities just outside the elusion region may be observed before the limiting value $h(t\rightarrow\infty)=$

Table 2
Values of SSR obtained by means of the EMG, GEMG, PMG1 and PMG2 functions at neutral mobile phases (pH 7) consisting of 1% and 2% isopropanol

ϕ (%)	Solute	EMG	GEMG	PMG1	PMG2
ED system					
1	5ht	503.1	39.4	37.5	36.0
1	5htp	261.7	8.1	51.9	13.3
1	3mt	356.8	47.8	8.3	10.8
1	mhpg	14.0	0.7	1.5	1.6
1	5hiaa	2.4	1.1	0.4	1.5
UV detector					
1	5ht	0.125	0.095	0.095	0.094
1	5htp	0.078	0.064	0.079	0.071
1	3mt	0.048	0.048	0.048	0.048
1	mhpg	0.044	0.044	0.040	0.039
1	5hiaa	0.046	0.045	0.041	0.041
ED system					
2	5ht	1378.6	90.0	19.5	27.7
2	5htp	588.4	8.9	58.6	61.0
2	3mt	418.3	92.5	7.4	8.9
2	mhpg	33.6	0.5	2.4	1.7
2	5hiaa	20.8	0.8	2.3	3.0
UV Detector					
2	5ht	0.130	0.024	0.019	0.015
2	5htp	0.058	0.008	0.039	0.039
2	3mt	0.025	0.016	0.015	0.020
2	mhpg	0.004	0.004	0.004	0.004
2	5hiaa	0.020	0.020	0.017	0.017

0 is attained (Fig. 1). For this reason we must redefine both these two functions by $h(t)=0$ outside the elusion region.

Taking into account the above observations we used the EMG function and the six-parameter PMG1 and PMG2 functions to deconvolute UV overlapped peaks, whereas for the corresponding peaks of the ED system we used the GEMG function and the eight-parameter PMG1 and PMG2 functions.

4.2. Reproducibility of the obtained results

The reproducibility of the experimental peaks and consequently the reproducibility of the results obtained from the fitting procedure is a necessary condition for an effective deconvolution of overlapped peaks. In order to examine this feature we have proceeded as follows: four peaks for each individual solute have been recorded by making four repetitive injections. Then the UV peaks were analysed by means of the EMG function and the ED peaks by means of the GEMG function. The percent error in the peak parameters was taken as a measure of the reproducibility of the obtained results and it was calculated from $100\Delta x/\langle x\rangle$, where $\langle x\rangle$ is the mean value of parameter $x(=A, \sigma, r)$ and Δx is the mean value of the quantity $|x-\langle x\rangle|$. Similarly the GEMG function was used for the calculation of the percent error in the area, A, of the ED peaks.

The results obtained by UV detection with 5% isopropanol are given in Table 3. The percent error in the area, A, of the ED peaks is 3.2 at pH 7 and 2.3 at pH 3. In the same table we have included the corresponding results obtained from simulated data. In particular, we have used the EMG function to reproduce the UV peaks by adding a Gaussian noise with standard deviation equal to ± 0.00002 AU. The Gaussian noise was generated in Excel by means of

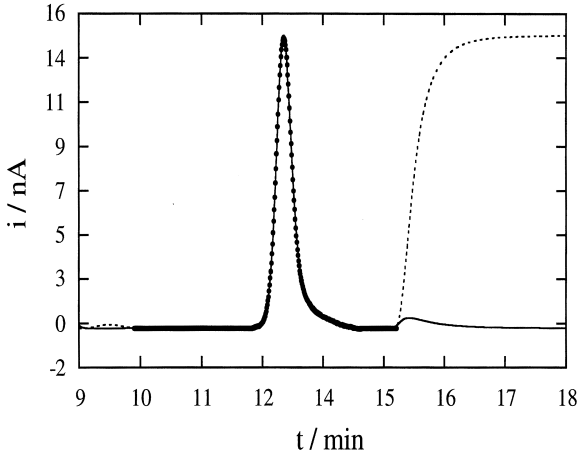

Fig. 1. PMG1 (- - -) and PMG2 (—) functions fitted to the experimental chromatogram (···) of 1.76 μg/ml mhpg in the presence of 1% isopropanol. Parameters used: $h_m = 14.903$, $t_m = 12.768$, $s_0 = 0.1867$, $s_1 = 0.0737$, $s_2 = 0.3284$, $s_3 = 0.0622$, $s_4 = -0.1137$, $s_5 = -0.02717$ (SSR = 1.46) for PMG1 and $h_m = 14.875$, $t_m = 12.776$, $s_0 = 0.1949$, $s_1 = 0.0917$, $s_2 = 0.377$, $s_3 = 0.1244$, $s_4 = -0.1422$, $s_5 = -0.04297$ (SSR = 1.70) for PMG2.

the *Random Number Generation* routine, following the sequence: *Tools→Data Analysis→Random Number Generation*. Then we choose the *Normal (Gaussian) Distribution* with zero mean and standard deviation equal to 0.00002. In this way the mean value of SSR of the experimental peaks was similar to that of the simulated data. The values of EMG parameters of the artificial data are given in Table 4. We observe that the errors in the simulated data are smaller than those in the real data. This shows that apart from the Gaussian noise there are other factors, like problems with the baseline and small fluctuations in temperature, that affect the accuracy of the obtained results [22].

The results of Table 3 correspond to data that have been obtained from the same solution. That is, the peaks of each solute have been obtained by four injections of the same solution. If we repeat this experiment by using new solutions for each injection and record one peak per day, then we observe a rather considerable increase in the errors of Table 3. Thus, the error in A is increased to 7.8% for the ED and to 5.6% for the UV peaks. The mean value in the error of σ is 4.1% and that of τ is 6.2%. These differentiations are attributed to problems with the baseline, errors in the preparation of solutions and fluctuations in temperature, as discussed in [22]. In addition, in case of ED peaks a passivation of the working electrode is possible.

4.3. Deconvolution of artificial peaks

The artificial data used in this section were taken from literature in order to compare the present method with the results of the immune algorithm and the genetic algorithm [1]. The first artificial six-

Table 3
Values of the percent error in the EMG parameters for UV peaks

Parameter	5% Isopropanol, pH 7		5% Isopropanol, pH 3	
	UV (experimental)	UV (simulated)	UV (experimental)	UV (simulated)
A	2.9	0.4	2.1	0.2
σ	0.8	0.4	0.5	0.3
r	4.6	1.8	3.6	1.3

Table 4
Values of the EMG parameters of the artificial data

Parameter	5% Isopropanol, pH 7					5% Isopropanol, pH 3		
	5ht	5htp	3mt	mhpg	5hiaa	5ht	5htp	m5ht
A	0.00072	0.001	0.00025	0.00017	0.015	0.00075	0.001	0.001
t_g	3.88	3.53	3.35	4.2	3.65	3.6	3.76	3.69
σ_g	0.044	0.04	0.038	0.05	0.04	0.043	0.048	0.045
τ	0.14	0.04	0.095	0.035	0.03	0.031	0.028	0.031

component chromatogram, chromatogram A, was created by means of the EMG function. The parameters used in the simulation of the six peaks of the chromatogram were the following. 0.2, 0.5, 0.3, 0.8, 0.5, 0.2 for A; 3.0, 3.8, 4.6, 5.3, 6.0, 6.6 for t_m; 1.5, 1.7, 1.9, 2.1, 2.3, 2.5 for σ and 1.5, 1.6, 1.7, 1.8, 1.8, 2.0 for τ. The time t was varied from 0 to 10 min by a step equal to 0.025. The final overlapping peak was obtained by the addition of the individual EMG peaks plus 3% random noise (Fig. 2A).

As we have already pointed out in the previous section, in chromatography we can know the peak characteristics of the individual solutes and this is attained by means of injection of standard solutions.

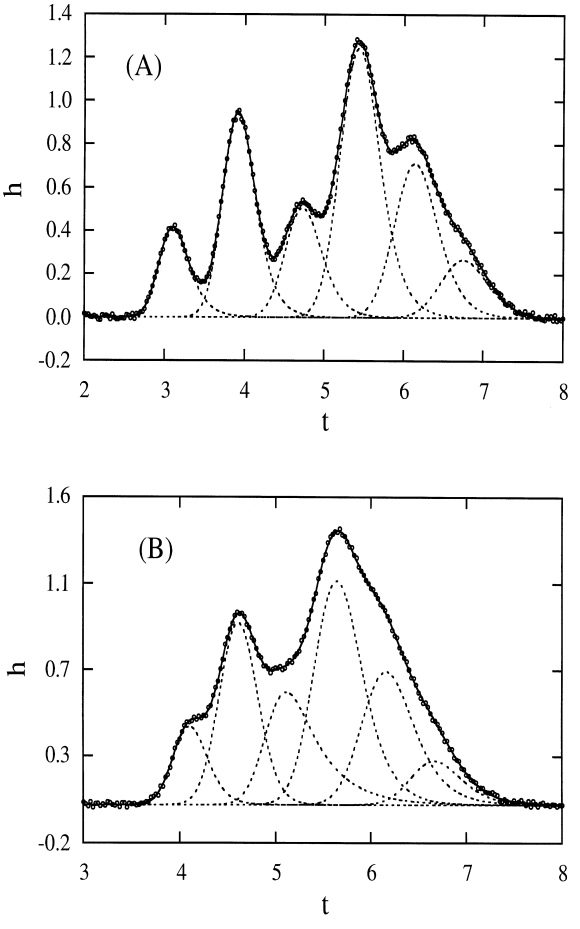

Fig. 2. Deconvolution of the simulated chromatograms A (A) and B (B). Points (◦) are simulated data, the solid line is the graph of the fitted EMG function and the broken lines correspond to the individual deconvoluted peaks.

Therefore from the peaks of the standard solutions we can obtain very good initial estimates for the fitting parameters t_m, σ and τ of the EMG function. When we use artificial data this information is obtained if we add 3% random noise to each individual peak and then fit the EMG function to this peak. Thus, the values of t_m, σ and τ obtained from the fitting of the EMG function to each individual peak were used as initial estimates in the FAP approach, whereas in the FSP approach the values of σ and τ were kept constant. For A we used arbitrary values aiming the calculated curve to be close to the artificial data.

It is well known that Solver as well as every other non-linear least-squares procedure based on the Levenberg–Marquardt algorithm depending on the initial estimates may converge to a local solution and not to the global minimum [9]. In order to avoid this situation we run Solver many times by adding a random noise to the initial estimates. In particular, the noise added to t_m was $\pm 5\%$ and that to the variables σ and τ in the FAP approach was $\pm 10\%$. The random noise was generated by the Uniform Distribution function of Excel as the Normal one with parameters between $-0.05t_m$ and $0.05t_m$ or from -0.1σ (-0.1τ) to 0.1σ (0.1τ). We found that the FSP approach always converges on the same solution, whereas the FAP approach may give more than one solution. However, all these solutions are characterised by SSR values that differ very little among themselves, in the fifth decimal point in our case. From these solutions we selected that with the minimum SSR value.

In Table 5 we present the original values of A, A(orig), the calculated ones from the fitting procedure values of A, A(calc), the relative error in the calculation of A, that is the quantity $100[A(\text{calc}) - A(\text{orig})]/A(\text{orig})$, the values of SSR of each fitting as well as the corresponding sum of the absolute values of residuals, i.e., the sum of absolute deviations (SAD). Note that the immune algorithm gives the value (average) SAD = 5.72 and the genetic algorithm the value SAD = 9.73 [1]. Note also that the added error in our calculations, i.e., the total sum of the random noise, was 6.25, whereas that in [1] was 5.08. We observe that Solver gives much better results than that of the immune and genetic algorithms.

Table 5
Adequacy of the method used for deconvoluting overlapping synthetic peaks

Solute	A(orig)	Chromatogram A				Chromatogram B			
		FSP		FAP		FSP		FAP	
		A(calc)	% Error	A(calc)	% Error	A(calc)	% Error	A(calc)	% Error
1	0.2	0.2041	2.05	0.2059	2.95	0.1976	−1.20	0.1845	−7.75
2	0.5	0.5005	0.10	0.5033	0.66	0.4990	−0.20	0.4762	−4.76
3	0.3	0.3001	0.03	0.3435	14.50	0.2986	−0.47	0.4248	41.6
4	0.8	0.8011	0.14	0.8034	0.42	0.7970	−0.38	0.7635	−4.56
5	0.5	0.4988	−0.24	0.4074	−18.52	0.5018	0.36	0.4943	−1.14
6	0.2	0.2005	0.25	0.2403	20.15	0.2035	1.75	0.1552	−22.4
SSR		0.0305		0.02995		0.0303		0.02998	
SAD		3.021		2.9779		3.003		2.9779	

However, the most important result comes from the comparison between the two approaches, the FAP and FSP. We observe that the results obtained from the FSP technique are much better than that of the FAP approach, despite the fact that the second approach leads to smaller values of SSR and therefore to a better fitting. Thus, contrary to what it is believed, the smaller value of SSR does not ensure the better deconvolution of an overlapped peak.

The above observations are further verified if we bring the individual peaks of the above peak closer to each other. Thus, all the EMG parameters of the above peak are kept constant except the time t_m which is changed to 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, respectively. The obtained overlapping peak (chromatogram B) with the addition of 3% random error is shown in Fig. 2B. The results of the deconvolution of chromatogram B are given in Table 5. It is seen that in principle Solver can deconvolute peaks with a high overlapping degree if the FSP approach is adopted. However, in practice this approach can be adopted only if the parameters that determine the peak shape of the individual solutes remain intact in overlapping peak. This assumption can be tested experimentally and this is done below, where we apply the Solver to deconvolute real multi-component peaks.

4.4. Deconvolution of experimental peaks

The results of the deconvolution of the experimental peaks are shown in Figs. 3–8 and in Tables 6–9. The area *A* in these tables is expressed in nA·min for the ED peaks and in mAU min for the UV peaks. Fig. 3 shows the deconvolution of the ED and UV chromatograms of 5ht, 5htp, 3mt, mhpg and 5hiaa in the presence of 1% isopropanol and pH 7 using the GEMG and EMG functions. In Fig. 4 the deconvolution of the ED chromatogram has been attempted by means of the PMG1 function using a sixth-order polynomial for *s*. Figs. 5 and 6 show the deconvolution of the ED and UV chromatograms of the above solutes in the presence of 2% and 5% isopropanol, respectively. It is seen that the increase of isopropanol content in the mobile phase brings the individual peaks closer to each other. In addition, we observe that the peaks of the chromatograms recorded by means of the ED system are much more asymmetric than those obtained by the UV detector. For this reason the peaks of the chromatograms of the ED system are more overlapped. When the acidity of the mobile phase is increased not only does retention time of the individual solutes vary but also the peaks asymmetry is decreased [23]. Thus, in the acidity mobile phase of pH 3 only the peaks of 5ht and 5htp are overlapped (Fig. 7). The addition of m5ht in the mixture of 5ht/5htp depicted in Fig. 8 has been done to create a three-component overlapping peak.

The researcher's ultimate purpose is to resolve an overlapping peak into individual peaks with areas as close as possible to those we would obtain if we chromatographed each solute separately. For this reason our criterion for the deconvolution of an

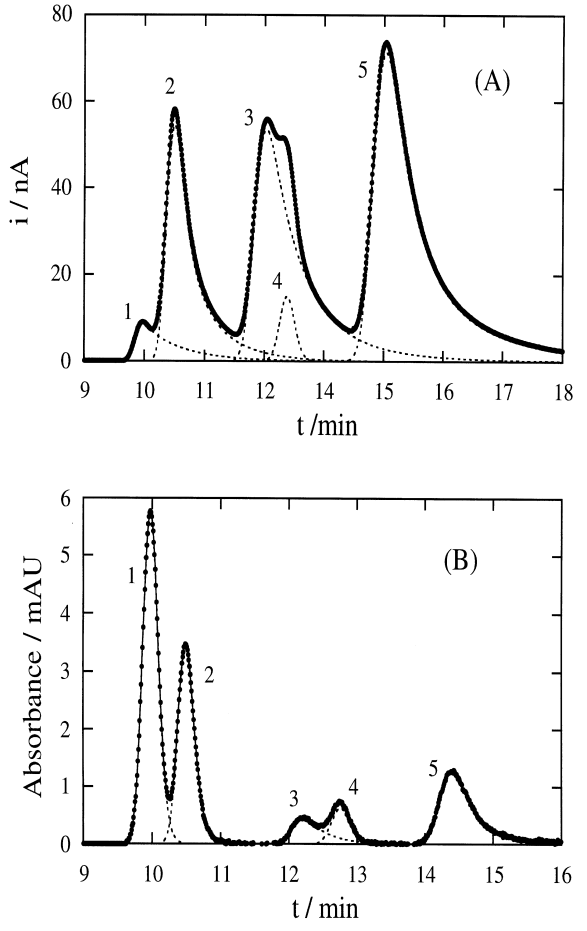

Fig. 3. Deconvolution of an ED (A) and a UV (B) chromatogram of a five-component mixture in the presence of 1% isopropanol and pH 7. Points (·) are experimental data, the solid line is the graph of the fitted GEMG (A) and EMG (B) function using the FSP approach and the broken lines correspond to the individual deconvoluted peaks: (1) 5hiaa, (2) 5htp, (3) 3mt, (4) mhpg and (5) 5ht. The standard mixture contains 2 μg/ml of each solute except mhpg the concentration of which is 1.76 μg/ml.

overlapped peak is not the value of SSR or SAD but the relative error in the calculation of the area of the individual peaks. These errors when we use for deconvolution the EMG, GEMG and PMG1 functions are given in Tables 6–9. The PMG2 function gives results that lie between those of the GEMG and PMG1 functions closer to the GEMG function.

We observe that when the FSP technique is used with the EMG or GEMG function, the deconvolution is very accurate even in cases of highly overlapped

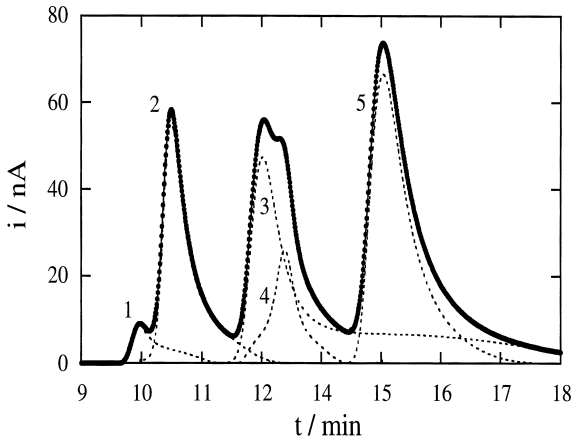

Fig. 4. As in Fig. 3A using the PMG1 function and the FAP approach for deconvolution.

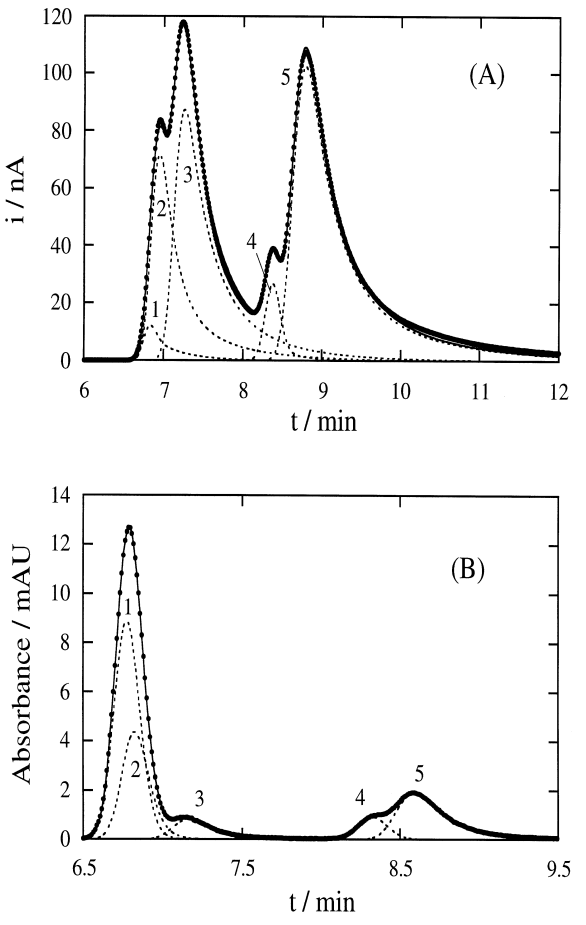

Fig. 5. As in Fig. 3 but in the presence of 2% isopropanol.

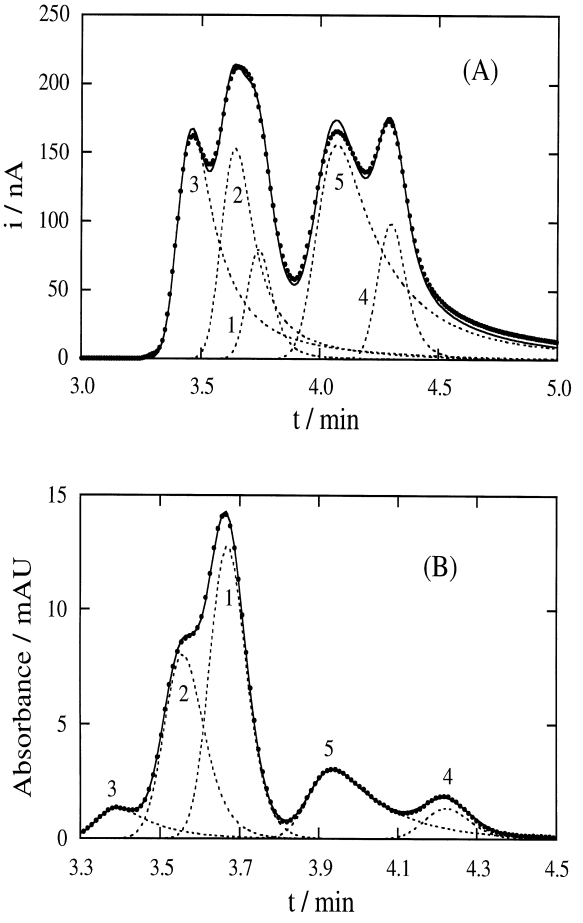

Fig. 6. As in Fig. 3 but in the presence of 5% isopropanol.

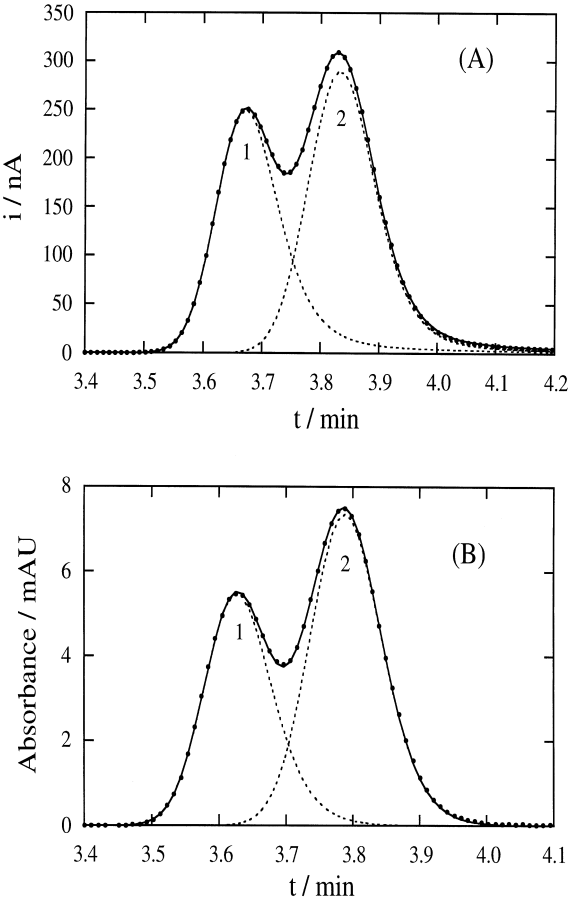

Fig. 7. Deconvolution of an ED (A) and a UV (B) chromatogram of a two-component mixture in the presence of 5% isopropanol and pH 3. Points (·) are experimental data, the solid line is the graph of the fitted GEMG (A) and EMG (B) function using the FSP approach and the broken lines correspond to the individual deconvoluted peaks: (1) 5ht and (2) 5htp.

peaks. The PMG1 function gives also good results, although in certain cases the error in the calculation of A is higher. On the contrary, the FAP technique gives poor results, especially when it is used with the PMG1 function. This is clearly depicted in Fig. 4 where the extracted individual peaks can hardly resemble chromatographic peaks (see for example peaks 3 and 4). The good results obtained from the FSP technique show that the parameters that determine the peak shape of the individual solutes remain unchanged in the multi-component unresolved peak.

It is quite interesting to focus our attention to the three-component system of 5ht, 5htp and m5ht in 5% isopropanol and pH 3 (Fig. 8). Despite the fact that the ED and UV chromatograms of this system exhibit highly overlapped peaks, their deconvolution with the FSP technique is satisfactory. In order to clarify whether this successful deconvolution is accidental or not, we proceeded to simulate the UV chromatogram, as in the case of the artificial data. Here, the noise we added was Gaussian noise of zero mean and standard deviation equal to 0.00002 AU. The values of σ and τ of the EMG parameters used for the artificial data are the same with those given in Table 4. For A and t_m we used the following values: $A=0.750$, $t_m=3.58$ min for 5ht, $A=1.001$, $t_m=3.79$ min for 5htp, and $A=0.999$, $t_m=3.65$ min for m5ht. The results we obtained are given in Table 10 and in

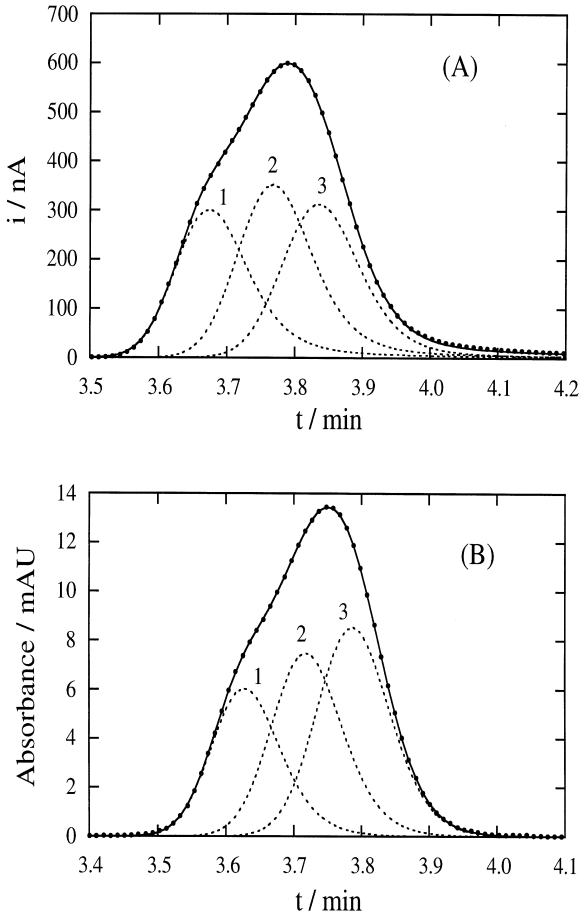

Fig. 8. Deconvolution of an ED (A) and a UV (B) chromatogram of a three-component mixture in the presence of 5% isopropanol and pH 3. Points (·) are experimental data, the solid line is the graph of the fitted GEMG (A) and EMG (B) function using the FSP approach and the broken lines correspond to the individual deconvoluted peaks: (1) 5ht, (2) m5ht and (3) 5htp.

Fig. 9. It is seen that the expected error in A is extremely small when we use the FSP approach, whereas the FAP approach gives higher but acceptable results. Thus, Solver can very effectively deconvolute the peaks of Fig. 8. The question raised is why the errors in A obtained from the deconvolution of the experimental peaks are systematically higher than those of the simulated data. As we have already pointed out, the shape of a chromatographic peak is not determined exclusively from the added noise but also from spurious factors discussed previously. These factors affect slightly but decisively the peak shape. Thus, if we increase the values of σ and r of 5htp by 1% and 4%, respectively, something which is within the experimental error, and use these values for the deconvolution of the simulated peak in the FSP approach, then we obtain the following results: % error in $A = -2.2$, 12.2 and -12.7 for 5ht, 5htp and m5ht, respectively. It is seen that the error increases considerably and becomes comparable to that detected experimentally. Therefore, for highly overlapped peaks the accuracy of the profile of the individual peaks is a basic prerequisite for an effective deconvolution.

5. Conclusions

The majority of the functions proposed to describe chromatographic peaks are not suitable for this purpose. The simple and the asymmetric Gaussian functions, the generalised exponential and the Lorentzian functions failed to describe our experimental data. The exponentially modified Gaussian (EMG) function was also problematic in describing peaks recorded by ED characterised by tailing behaviour in some cases.

From the functions of the literature only the polynomial modified Gaussian (PMG1) can describe almost every peak. However, it is mathematically incorrect unless it is redefined properly, since it does not tend to zero when t tends to infinity. In addition, its high flexibility due to the use of polynomial s may become a boomerang when it is used for deconvolution of overlapped peaks, because if the overlapping region is long enough, the resolved individual peaks may not resemble single chromatographic peaks of pure compounds (Fig. 4).

The proposed generalised exponentially modified Gaussian function (GEMG) extends satisfactorily the EMG function in cases of very asymmetric peaks. The proposed modification of the PMG1 function has some advantages but in general it should be handled with care, as the original PMG1 function.

The deconvolution of an overlapping peak can be done with high accuracy using a non-linear least-squares procedure, like Solver. However, this target is attained only if we know the actual number of components of the unresolved peak and we use as fitted parameters the position of the peak maximum

Table 6
Adequacy of the method and the functions used for deconvoluting overlapping peaks at mobile phases consisting of 1% isopropanol and pH 7

Solute	*A*(orig)	Method/function							
		FSP/EMG or GEMG		FSP/PMG1		FAP/EMG or GEMG		FAP/PMG1	
		A(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error
UV detector									
5ht	0.760	0.779	2.50	0.804	5.79	0.787	3.55	0.854	12.37
5htp	1.030	1.059	2.82	1.060	2.91	1.054	2.33	1.071	3.98
3mt	0.240	0.250	4.17	0.245	2.08	0.261	8.75	0.230	−4.17
mhpg	0.211	0.213	0.95	0.231	9.48	0.207	−1.90	0.234	10.90
5hiaa	1.590	1.610	1.26	1.624	2.14	1.634	2.77	1.622	2.01
SSR		0.209		0.394		0.174		0.243	
ED system									
5ht	77.70	79.80	2.70	76.85	−1.09	82.22	5.82	62.85	−19.11
5htp	35.60	37.64	5.73	37.07	4.13	35.63	0.08	36.00	1.12
3mt	60.61	62.20	2.62	57.90	−4.47	56.70	−6.45	62.15	2.54
mhpg	5.86	6.11	4.27	5.54	−5.46	5.35	−8.70	16.91	188.4
5hiaa	5.95	5.85	−1.68	5.85	−1.68	7.87	32.27	5.95	0.00
SSR		191.6		212.3		43.4		10.9	

Table 7
Adequacy of the method and the functions used for deconvoluting overlapping peaks at mobile phases consisting of 2% isopropanol and pH 7

Solute	*A*(orig)	Method/function							
		FSP/EMG or GEMG		FSP/PMG1		FAP/EMG or GEMG		FAP/PMG1	
		A(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error
UV detector									
5ht	0.760	0.767	0.92	0.787	3.55	0.702	−7.63	0.75	−1.32
5htp	1.027	0.981	−4.48	0.964	−6.13	1.231	19.86	1.072	4.38
3mt	0.240	0.260	8.33	0.274	14.17	0.282	17.50	0.312	30.00
mhpg	0.210	0.211	0.48	0.213	1.43	0.28	33.33	0.232	10.48
5hiaa	1.585	1.680	5.99	1.770	5.99	1.476	−6.88	1.624	2.46
SSR		0.155		0.176		0.105		0.031	
ED system									
5ht	77.5	82.7	6.71	80.35	3.68	84.22	8.67	82.66	6.67
5htp	35.4	35.41	0.03	34.57	−2.34	16.66	−52.94	25.51	−27.94
3mt	53.3	55.32	3.79	53.48	0.34	75.70	42.03	62.00	16.32
mhpg	7.71	7.33	−4.93	7.60	−1.43	5.73	−25.68	7.14	−7.39
5hiaa	5.10	4.73	−7.25	6.52	27.84	7.24	41.96	7.92	55.29
SSR		738.0		1067.6		23.5		1.16	

Table 8
Adequacy of the method and the functions used for deconvoluting overlapping peaks at mobile phases consisting of 5% isopropanol and pH 7

Solute	A(orig)	Method/function							
		FSP/EMG or GEMG		FSP/PMG1		FAP/EMG or GEMG		FAP/PMG1	
		A(calc)	% Error	A(calc)	% Error	A(calc)	% Error	A(calc)	% Error
UV detector									
5ht	0.750	0.731	−2.53	0.713	−4.93	0.726	−3.20	0.600	−20.0
5htp	1.030	1.040	0.97	1.062	3.11	0.840	−18.45	0.914	−11.26
3mt	0.240	0.240	0.00	0.246	2.50	0.226	−5.83	0.452	88.33
mhpg	0.210	0.200	−4.76	0.196	−6.67	0.185	−11.90	0.171	−18.57
5hiaa	1.542	1.490	−3.37	1.480	−4.02	1.700	10.25	1.543	0.06
SSR		0.363		0.334		0.077		0.052	
ED system									
5ht	70.40	65.15	−7.46	63.20	−10.23	76.11	8.11	64.82	−7.93
5htp	32.51	33.34	2.55	33.87	4.18	13.54	−58.35	22.14	−31.90
3mt	44.50	43.35	−2.58	43.40	−2.47	49.26	10.70	57.22	28.58
mhpg	17.50	15.45	−11.71	15.45	−11.71	14.22	−18.74	14.33	−18.11
5hiaa	12.40	11.45	−7.66	10.63	−14.27	20.62	66.29	14.52	17.10
SSR		2802.0		3246.0		84.5		10.6	

t_m and the peak area A or the peak height h_m of every component of the peak. If this prerequisite is adopted, the choice of the function used (EMG, GEMG, PMG1, PMG2) to describe the peak profile does not play an important role. It is evident that if an impurity or an unknown solute contributes to the overlapped peak, our method cannot be applied.

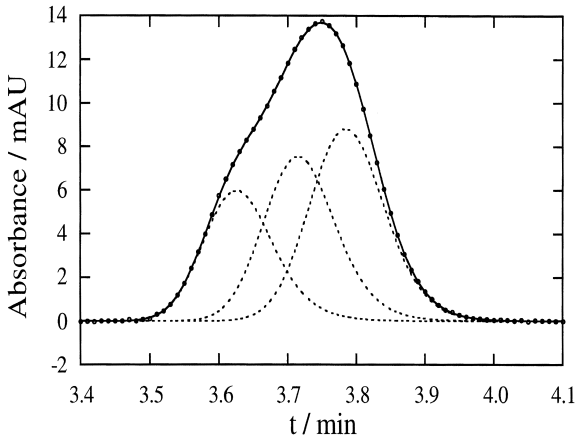

Fig. 9. Deconvolution of the simulated chromatogram of Fig. 8B. Points (·) are simulated data, the solid line is the graph of the fitted EMG function and the broken lines correspond to the individual deconvoluted peaks.

If we use as fitted parameters all the parameters that describe a peak, then Solver leads to solutions with considerable smaller values of SSR or SAD. However, these solutions are usually erroneous, especially if the PMG1 function is used. Thus, the smaller value of SSR or SAD does not ensure the better deconvolution of an overlapping peak.

In general, if we take into account the above observations, Solver performs with high accuracy giving much better results than those of modern methods, like the immune and genetic algorithms. Thus, Solver is a convenient tool for the resolution of overlapping chromatograms. The practical interest of the proposed approach concerns many cases such as when: (a) it is not feasible to resolve a mixture due to the nature of eluting species, (b) some resolution is sacrificed to decrease the analysis time, and (c) it is necessary to use chromatographic conditions which contribute to peaks asymmetry and consequently to incomplete resolution between adjacent peaks.

Table 9
Adequacy of the method and the functions used for deconvoluting overlapping peaks at mobile phases consisting of 5% isopropanol and pH 3

Solute	*A*(orig)	Method/function							
		FSP/EMG or GEMG		FSP/PMG1		FAP/EMG or GEMG		FAP/PMG1	
		A(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error	*A*(calc)	% Error
UV detector									
5ht	0.750	0.775	3.33	0.776	3.47	0.790	5.33	0.450	−40.0
5htp	1.031	1.120	8.63	1.155	12.03	1.073	4.07	1.392	35.0
m5ht	1.050	0.986	−6.10	0.986	−6.10	1.060	0.95	1.084	3.24
SSR		0.036		0.061		0.034		0.019	
ED system									
5ht	39.20	44.31	13.04	44.23	12.83	37.73	−3.75	43.95	12.12
5htp	45.64	47.70	4.51	46.76	2.45	50.42	10.47	50.26	10.12
m5ht	54.42	52.52	−3.84	52.72	−3.48	56.87	4.12	50.80	−6.99
SSR		234.5		335.0		24.3		26.5	
UV detector									
5ht	0.720	0.700	−2.78	0.700	−2.78	0.725	0.69	0.675	−6.25
5htp	1.010	1.001	−0.89	0.990	−1.98	0.980	−2.97	1.011	0.10
SSR		0.056		0.058		0.023		0.077	
ED system									
5ht	38.20	36.65	−4.06	36.95	−3.27	34.87	−8.72	43.21	13.12
5htp	45.60	44.16	−3.16	43.95	−3.62	45.51	−0.20	37.08	−18.68
SSR		60.9		136.2		18.3		8.0	

Table 10
Adequacy of the method used for deconvoluting overlapping synthetic peaks that simulate the system of 5ht, 5htp and m5ht in 5% isopropanol and pH 3

Solute	*A*(orig)	Method/function			
		FSP/EMG		FAP/EMG	
		A(calc)	% Error	*A*(calc)	% Error
5ht	0.750	0.748	−0.27	0.763	1.73
5htp	1.001	1.006	0.50	1.024	2.30
m5ht	0.999	0.994	−0.50	0.957	−4.20
SSR		0.083		0.082	

References

[1] X. Shao, Z. Chen, X. Lin, Chemometr. Intell. Lab. Syst. 50 (2000) 91.
[2] H. Miao, M. Yu, S. Hu, J. Chromatogr. A 749 (1996) 5.
[3] J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coque, J.J. Baeza-Baeza, Anal. Chim. Acta 348 (1997) 187.
[4] S. Walsh, D. Diamond, Talanta 42 (1995) 561.
[5] J.O. Machuca-Herrera, J. Chem. Educ. 74 (1997) 448.
[6] E.J. Billo, Excel for Chemists, A Comprehensive Guide, Wiley–VCH, 1997.
[7] D.C. Harris, J. Chem. Educ. 75 (1998) 119.
[8] R. DeLevie, J. Chem. Educ 76 (1999) 1594.
[9] P. Nikitas, A. Pappa-Louisi, Chromatographia 52 (2000) 477.
[10] E. Grushka, Anal. Chem. 44 (1972) 1733.
[11] W.E. Barber, P.W. Carr, Anal. Chem. 53 (1981) 1939.
[12] J.P. Foley, J.G. Dorsey, Anal. Chem. 55 (1983) 730.
[13] D. Hanggi, P.W. Carr, Anal. Chem. 57 (1985) 2395.
[14] R.A. Vaidya, R.D. Hester, J. Chromatogr. 287 (1984) 231.
[15] J.P. Foley, Anal. Chem. 59 (1987) 1984.
[16] S. Sekulic, P.R. Haddad, J. Chromatogr. 459 (1988) 65.
[17] S. Carda-Broch, J.R. Torres-Lapasio, M.C. Garcia-Alvarez-Coqueza, Anal. Chim. Acta 396 (1999) 61.
[18] K. Levenberg, Quart. Appl. Math. 2 (1944) 164.
[19] D.W. Marquardt, J. SIAM 11 (1963) 41.

[20] W.H. Press, B.P. Flanerry, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, 1986.
[21] G.M. Cao, J. Chromatogr. A 746 (1996) 161.
[22] N. Dyson, Chromatographic Integration Methods, Royal Society of Chemistry, London, 1998.
[23] A. Pappa-Louisi, F. Zougrou, Chromatographia 44 (1997) 348.