ELSEVIER

Journal of Chromatography A, 952 (2002) 63–70

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Comparison of the capability of peak functions in describing real chromatographic peaks

Jianwei Li*

Analytical Research and Development, 3M Pharmaceuticals, 3M Center, Building 235-BE-45, St. Paul, MN 55144, USA

Received 13 November 2001; received in revised form 23 January 2002; accepted 24 January 2002

Abstract

This paper describes the results of a comparison of four peak functions in describing real chromatographic peaks. They are the empirically transformed Gaussian, polynomial modified Gaussian, generalized exponentially modified Gaussian and hybrid function of Gaussian and truncated exponential functions. Real chromatographic peaks of different shapes (fronting, symmetric, and tailing) are obtained by various separation conditions of reversed-phase liquid chromatography. They are then fitted to the peak functions via the Marquardt–Levenberg algorithm, a nonlinear least-squares curve-fitting procedure, by Microsoft Solver. The qualities of the fits are evaluated by the sum of the squares of the residuals. It is concluded in the study that the empirically transformed Gaussian function offers the highest flexibility (best fits) to all shapes of chromatographic peaks, including extremely asymmetric tailing peaks with a peak asymmetry of up to 8. The flexibility of this function should improve our ability to process chromatographic peaks such as deconvolution of overlapped peaks and smoothing noisy peaks for the determination of statistical moments.

Keywords: Peak functions; Peak modeling; Curve fitting; Moment analysis; Deconvolution

1. Introduction

Although the description of chromatographic peaks by mathematical models is of considerable importance in chromatographic science, the search for flexible peak functions, both theoretically and empirically, has been challenged by a significant variation in the shape of chromatographic peaks. There are at least two major applications of peak functions in chromatographic data processing, and they include the deconvolution of overlapped peaks and smoothing experimental peaks for the determination of the statistical moments [1,2]. The accurate determination of statistical moments is critical to evaluate the thermodynamic properties and mass transfer kinetics of separation processes [3–10].

However, the application of the curve-fitting approach for either deconvolution or smoothing noises depends on the capability of these functions to describe real chromatographic peaks. It is highly desirable that the mathematical models can describe the peaks perfectly. In other words, the peak functions should possess sufficient flexibility to fit peaks of different shapes. The most popular model is the exponentially modified Gaussian (EMG) function [11]. However, the flexibility of this function is rather limited.

The difficulty in obtaining, by mathematical means, an explicit equation describing the shape of a

*Tel.: +1-651-737-0468; fax: +1-651-737-7918.
E-mail address: jli7@mmm.com (J. Li).

0021-9673/02/$ – see front matter
PII: S0021-9673(02)00090-0

chromatographic peak has motivated the construction of equations that do not have a rigorous physicochemical foundation but acceptably fit experimental peaks. In a previous publication, a flexible peak function (called empirically transformed Gaussian function, ETG) was proposed based on the decomposition of Gaussian function into two separate leading and trailing functions [12]. The two functions are then empirically combined and modified to provide a flexible model. The capability and flexibility of the function is due to the fact that the leading edge of the model is weakly related to the trailing edge and vice versa. This function has been shown to perfectly describe the EMG, Giddings, Haarhoff–Van der Linde, Poisson, log-normal, statistical, nonlinear chromatography functions, and Edgeworth–Cramér series.

Since then several other newer functions have also been published. They are the polynomial modified Gaussian (PMG), generalized exponentially modified Gaussian (GEMG) function [13,14] and a hybrid function of Gaussian and truncated exponential functions (EGH) [15]. All these functions are empirical in nature, and they were reported to offer better fits than the popular EMG model. Moreover, these functions have been used to deconvolute overlapped peaks and to fit very asymmetric tailing peaks. It should be pointed out that ETG function is designed to fit symmetric, fronting and tailing peaks, while other three functions are essentially designed to fit tailing peaks.

The execution of the nonlinear least-squares curve-fitting procedure used to fit chromatographic peaks can be considerably simplified now by commercial spreadsheet programs such as Microsoft Solver [14]. No computer programming experience is required. It is expected that the capability of commercial software helps greatly the use of peak functions to process chromatographic peaks.

The goal of the study is to compare the capability of recently published peak functions to describe real chromatographic peaks with an emphasis on ETG function. Chromatographic peaks with different shapes (fronting, symmetric, and tailing) are obtained under various separation conditions of reversed-phase liquid chromatography (RPLC). These peaks are fitted by the peak functions to evaluate the capability of each model based on the quality of the fits.

2. Mathematical description

As mentioned in the Introduction, four major peak functions have been published in the last several years. These functions were shown to offer better fits to real chromatographic peaks than the preceding peak models (for example, the EMG function). These four functions are compared in the study, and described below.

2.1. Empirically transformed Gaussian function (ETG)

This function is essentially an empirical model. It is based on the decomposition of Gaussian function into leading and trailing edge functions [12]. The two functions are weakly combined and subsequently modified (exponentially). Although several combinations have been suggested, the most versatile and useful one is given by:

$$f(t) = \frac{H''}{\{1+\lambda_L \exp[k_L(t_L-t)]\}^{(t_L/t)^{\alpha}} + \{1+\lambda_T \exp[k_T(t-t_T)]\}^{(t/t_T)^{\beta}} - 1} \quad (1)$$

where $f(t)$ denotes peak function; t denotes time; H'' is related to peak amplitude; λ_L and λ_T are pre-exponential parameters, k_L and k_T are the parameters related to the speeds of the rise and fall of the leading and trailing edges, respectively; t_L and t_T are the inflection times of the leading and trailing edges, respectively; and α and β are the parameters to further modify the shapes of the leading and trailing edges, respectively. The design of the function offers a weak link between leading and trailing parts of a peak, and therefore, improves the ''rigidity'' of the most peak functions. Although seven parameters are used in the function, it converges rapidly. The selection of t_L and t_T does not have to be accurate, and can be replaced by the times at half height.

2.2. Polynomial modified Gaussian function (PMG)

The function was also empirically proposed based on a variable standard deviation in Gaussian function [13,14]. The more stable version of the model (PMG2) is proposed as:

$$f(t) = \frac{H_m \sigma_0}{\sigma} \exp\left[-\left(\frac{t - t_m}{\sigma}\right)^2 \right] \quad (2)$$

where H_m is an amplitude parameter; t_m is the time of peak center; σ_0 is a constant standard deviation; and σ is the polynomial standard deviation defined as below:

$$\sigma = \sigma_0 + \sigma_1(t - t_m) + \sigma_2(t - t_m)^2 + \sigma_3(t - t_m)^3 + \sigma_4(t - t_m)^4$$

where σ_1, σ_2, σ_3, and σ_4 are fitting parameters. Five terms in the variable standard deviation are used in the study. However, the number of terms can be adjusted.

2.3. Generalized exponentially modified Gaussian function (GEMG)

This function is obtained by convoluting a Gaussian function with the resultant of two exponential functions of different time constants [14]. Such a function may be expressed as:

$$f(t) = \frac{A}{1 + b}\left[\frac{\exp(q_1) I_1(z_1)}{\tau_1} + \frac{\exp(q_2) I_2(z_2)}{\tau_2} \right] \quad (3)$$

where A is peak area; b is the percentage contribution of the second exponential function; τ_1 and τ_2 are the exponential decays; and q_i, I_i, and z_i ($i = 1$ and 2) are given by:

$$q_i = \frac{\sigma^2}{2\tau^2} - \frac{t - t_m}{\tau_i}$$

$$I_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} \exp\left(-\frac{x^2}{2} \right) dx$$

$$z_i = \frac{t - t_m}{\sigma} - \frac{\sigma}{\tau_i}$$

Programming this function into Microsoft Solver has been described elsewhere [14].

2.4. Hybrid of Gaussian and truncated exponential functions (EGH)

This function is obtained by combination of a Gaussian and truncated exponential functions [15]:

$$f(t) = \begin{cases} H \exp\left(\dfrac{-(t - t_m)^2}{2\sigma^2 + \tau(t - t_m)} \right), & 2\sigma^2 + \tau(t - t_m) > 0 \\ 0, & 2\sigma^2 + \tau(t - t_m) \le 0 \end{cases} \quad (4)$$

where H is the peak height; σ is the standard deviation of the parent Gaussian peak; and τ is the time constant of the precursor exponential decay. This function is the simplest (four parameters only) among the four models.

All four models have been shown to fit asymmetric peaks well. Table 1 summarizes the fitting parameters for each model.

Table 1
A summary of peak function fitting parameters

Model	No. of fitting parameters	Fitting parameters	Remarks on fitting
ETG	7	H'', λ_l, λ_t, k_l, k_t, n_l, n_t	Easy programming and rapid convergence
PMG2[a]	7	H_m, t_m, σ_0, σ_1, σ_2, σ_3, σ_4	Convergence can be slow due to rational nature of the function
GEMG[a]	6	A, b, t_m, σ, τ_1, τ_2	The most difficult function to fit due to the error function
EGH[a]	4	H_m, t_m, σ, τ	Fastest convergence due to the simplicity and number of parameters. Versatility may suffer

[a] The peak center (t_m) is treated as a parameter in the study.

3. Experimental

3.1. Chromatographic instrumentation

All experiments were performed on a HP 1090 chromatograph equipped with a ternary pump, autosampler, and diode-array detector (Agilent Technologies, Wilmington, DE, USA). A computer-based workstation (HP Chemstation) was used not only to control the instrumentation, but also to collect chromatographic data. The chromatographic data exported via CSV format can be directly loaded into Microsoft Excel.

3.2. Separation conditions and samples

Two columns were used to perform the separations. They were Inertsil ODS(3) (150×4.6 mm, 5 μm) (Metachem Technologies, Torrance, CA, USA) and Zorbax Eclipse columns (150×4.6 mm, 5 μm) (Agilent Technologies). The mobile phases were acetonitrile (ACN)–water. The separation conditions for each sample are summarized in Table 2. Two

Table 2
Separation conditions for the samples used in the study[a]

Peak shape	Sample	Separation conditions					
		Column	Column temperature (°C)	Mobile phase composition (v/v)	Flow rate (ml/min)	Injection volume (μl)	Detection wavelength[f] (nm)
Symmetric peaks	Benzophenone[b]	Eclipse ODS	Ambient	40% ACN	1	15	220
	Butyl paraben[b]			40% ACN		15	240
	4-Chlorophenol[b]			45% ACN		10	250
	Ethyl benzoate[b]			40% ACN		15	260
	4-Phenyl phenol[b]			40% ACN		15	270
	Propyl paraben[b]			40% ACN		15	280
	2,4,6-Trichlorophenol[b]			45% ACN		10	290
Fronting peaks	Ethyl paraben[b]			40% ACN		15	
	4-Nitrophenol[b]			40% ACN		15	
Tailing peaks	4-*tert*.-Butyl catechol[c]	Eclipse ODS	Ambient	40% ACN	1	15	290
	Antidepressants [16][d] (amitriptyline nortriptyline)	PBD-Zirconia	40	40% ACN (20 m*M* NH_4F)	1	1	254
	Chlorophenols [17][e] (2,4,6-ttichlorophenol, 2,4,5-trichlorophenol, pentachlorophenol)	PBD-Zirconia	40	32% ACN (20 m*M* NH_4F)	1	1	254
	Phenol/ethyl paraben[e]	Inertsil ODS(3)	Ambient	40% ACN	2	20	280
	4-Ethylpyridiene[d]	Eclipse ODS	Ambient	10% ACN (0.025% TFA)	1	5	220 240
	2,6-Dimethylaniline[d]	Eclipse ODS	Ambient	10% ACN (0.025% TFA)	1	5	250 260 270 280 290

[a] This table summarizes the separation conditions for the samples.
[b] The sample concentration~50 μg/ml.
[c] The sample concentration~5 μg/ml.
[d] The sample concentration~10 μg/ml.
[e] The sample concentration~1 mg/ml. Phenol peak belongs to the symmetric group.
[f] Only selected data are used in the study.

types of samples (neutral and amine) were used to provide different peak shapes. All samples were prepared by ACN and their concentrations were indicated in the table. All chemicals are obtained from Aldrich (Milwaukee, WI, USA). For the separation of aromatic amines (4-ethylpyridine and 2,6-dimethylaniline), the mobile phase contains 0.025% (v/v) trifluoroacetic acid (TFA). For several separations, multiple wavelength detection (220–290 nm) was used to obtain peaks with different S/N ratios, although not all the data are processed in the study. Table 2 also shows the separations of antidepressants and chlorophenols on polybutadiene coated zirconia (PBD-Zirconia) column (100×4.6 mm, 5.6% carbon) with ACN–water mobile phases. The two samples are ionizable and their peaks exhibited a significant tailing. These data have been published already [16,17], and are also used in the study. Table 3 shows the asymmetry and USP tailing factors of the peaks. The peaks are approximately categorized into three groups: symmetric, fronting and tailing based on their asymmetry at 5% peak height. The three types of peaks are used to evaluate the capability of each model to fit real chromatographic peaks.

3.3. Fitting chromatographic peaks

Before fitting the mathematical models to real chromatographic peaks, the baselines are corrected by subtracting the linear interpolation in regions before and after the elution of the peaks. The

Table 3
Comparison of curve-fitting results by different peak functions

Peak shape	Solute	Peak characteristics		Quality of fits (SSR)[a]				Remarks
		Asymmetry (5% peak height)	USP tailing factor	ETG	PMG2	GEMG	EGH	
Symmetric	Benzophenone	0.98	0.99	1.5	39.7 (26)	11.5 (8)	4.9 (3)	
	Butyl paraben	0.88	0.94	3.0	44.5 (15)	136 (45)	22.6 (8)	
	4-Chlorophenol	1.05	1.03	62.7	2029 (32)	128.9 (2)	150.5 (2)	
	Ethyl benzoate	0.98	0.99	0.6	400 (667)	52.0 (87)	25.3 (42)	220 nm, Fig. 1
	4-Phenylphenol	0.70	0.85	8.5	106 (12)	2540 (299)	101 (12)	
	Propyl paraben	0.72	0.86	8.3	36.4 (4)	1270 (153)	65.4 (8)	
	2,4,6-Trichlorophenol	1.05	1.03	0.2	20.5 (103)	0.2 (1)	0.5 (3)	
Fronting	Ethyl paraben	0.60	0.80	46.7	805 (17)	4408 (94)	423 (9)	220 nm, Fig. 2
	4-Nitrophenol	0.52	0.76	191	498 (3)	12 575 (66)	2246 (12)	
Tailing	4-*tert*.-Butyl catechol	3.03	2.02	0.1	1.7 (17)	0.2 (2)	0.7 (7)	290 nm, Fig. 3
	Amitriptyline	1.48	1.24	1.2	61.6 (51)	13.9 (12)	18.7 (16)	Antidepressants, 254 nm, Fig. 4
	Nortriptyline	1.92	1.46	5.6	91.4 (16)	28.4 (5)	95.3 (17)	
	2,4,6-Trichlorophenol	2.33	1.67	8.5	9.4 (1)	2.9 (0.3)	37.5 (4)	Chlorophenols[b], 254 nm, Fig. 5
	2,4,5-Trichlorophenol	1.48	1.24	6.5	48.5 (7)	12.1 (2)	12.1 (2)	
	Pentachlorophenol	3.28	2.14					
	Phenol	0.76	0.88	5772	11 184 (2)	24 270 (4)	12 122 (2)	Phenol peak
	Ethyl paraben	2.44	1.72	987	231 448 (234)	3341 (3)	172 089 (174)	is symmetric
	4-Ethylpyridine	3.33	2.16	38.1	1515 (40)	62.0 (2)	1109 (29)	220 nm, Fig. 6
	2,6-Dimethylaniline	7.68	4.34	39.8	47 849 (1202)	23 097 (580)	11 049 (278)	220 nm, Fig. 7
Average SSR ratio[c]				1	136	76	35	

[a] The SSR values are used to compare the fit quality by each function. The values in the parentheses indicate the SSR ratios relative to the ETG model.

[b] The last two peaks (2,4,5-trichlorophenol and pentachlorophenol) are fitted together.

[c] This is computed by (1) obtaining the ratio relative to ETG model; and (2) averaging the ratios.

integration range is wide enough, as confirmed by the first derivative of the peak. Microsoft Excel is used for the baseline treatment and all other manipulations of the experimental data. Furthermore, the Microsoft Solver, a very powerful routine based on the Marquardt–Levenberg algorithm, is used for the nonlinear least-squares curve-fitting. The procedures for the fitting were based on a process that has been described elsewhere [14]. The best fits are obtained by minimizing the sum of squares of residuals (SSR). The residual is the vertical (response) deviation of the estimated values from the data values.

The initial estimates of parameters are very important to the speed of the fitting. To facilitate the selection of initial parameter values, a real peak, mathematical function, and the residual are plotted together. By manually adjusting the initial values, the shape of the function can be visualized and compared to the peak. When there is a proper match between the peak and function shape, the Microsoft Solver is then started. The curve-fitting process is usually very fast, and no divergence was observed. Sometimes, after one sequence of iterations based on the default Solver convergence options, SSR is still not minimized (local minimum is reached). Therefore, Microsoft Solver is repeated at least three times for each fit until there is no change in SSR. The final value of SSR is used to evaluate the qualities of the fits. The fitting results are summarized in Table 3.

4. Results and discussion

4.1. Symmetric peaks

Eight symmetric peaks are used for comparison in the study, and the qualities of the fits in terms of SSR by each model are shown in Table 3. Fig. 1 graphically show the qualities of the fits for ethyl benzoate as an example. It can be seen from the figure and Table 3 that the ETG model fits the symmetric peaks almost perfectly (the best), as evidenced by the smallest residuals (see plots on the right). The range of SSR ratios is from 4 to 667 with an average of 108 for PMG2; from 1 to 299 with an average of 75 for GEMG; and from 2 to 42 with an average of 10 for EGH, respectively. Clearly, the fit

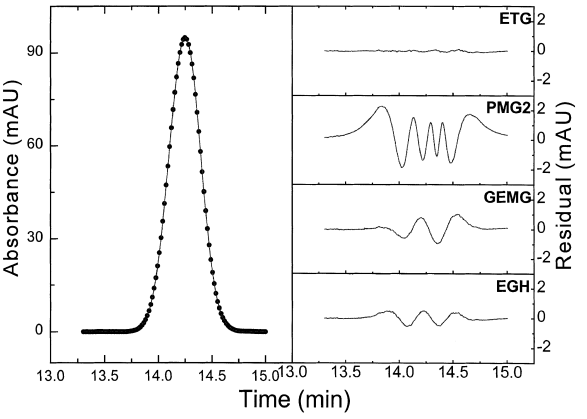

Fig. 1. Symmetric ethyl benzoate peak and the fitting results. See Table 2 for the separation conditions. The left graph shows the peak (solid circle, every third point) and fit (solid line) by ETG function. The right graph shows the fit residuals by the four peak functions. The residual is the vertical (response) deviation of the estimated values from the data values.

qualities can be arranged as follows: ETG>EGH>GEMG>PMG2.

4.2. Fronting peaks

Fig. 2 illustrates the fitting results for ethyl paraben as an example for the fronting peaks. It can be seen in Fig. 2 and Table 3 that the ETG function fits the fronting peaks nicely, and the qualities of the fits are better than other three models. The average SSR ratios are 10, 80, and 10 for PMG2, GEMG, and EGH, respectively. The results are consistent

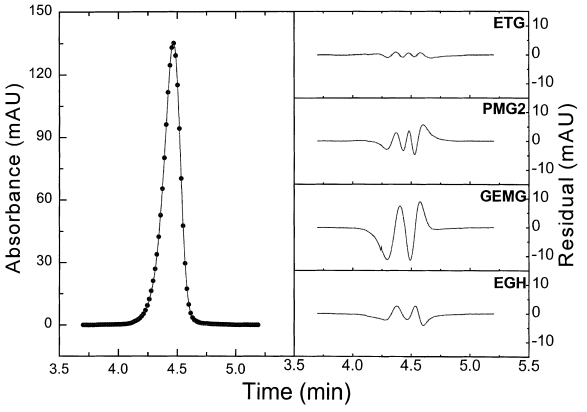

Fig. 2. Fronting ethyl paraben peak and the fitting results. All descriptions as in Fig. 1.

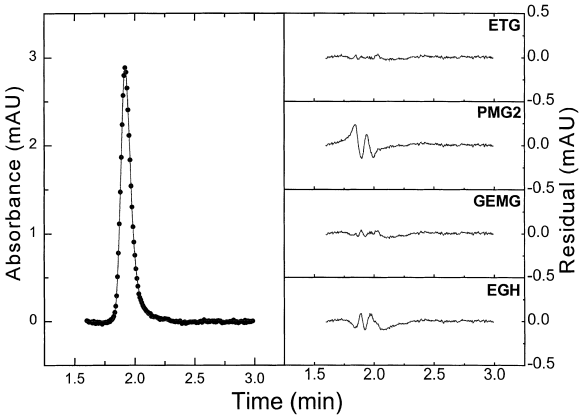

Fig. 3. Asymmetric tailing 4-*tert*.-butyl catechol peak and the fitting results. All descriptions as in Fig. 1.

with the design of ETG function (to fit both fronting and tailing peaks).

4.3. Tailing peaks

In most of chromatographic applications, peak tailing is the most common phenomenon due to the extra column broadening and secondary interaction with the stationary phase. Most of the peak functions focus on this type of peak shapes. Figs. 3–7 show the fit results for 4-*tert*.-butyl catechol, antidepressants (amitriptyline, first peak and nortriptyline, the second peak), chlorophenols, ethyl paraben, 4-

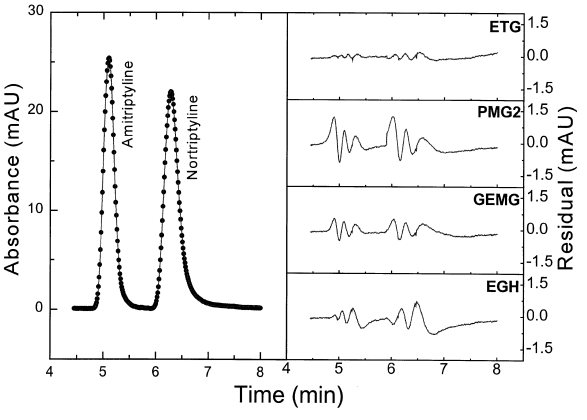

Fig. 4. Asymmetric tailing antidepressant peaks and the fitting results. The first peak is amitriptyline, and the second peak is nortriptyline. The two peaks are fitted independently. All other descriptions as in Fig. 1.

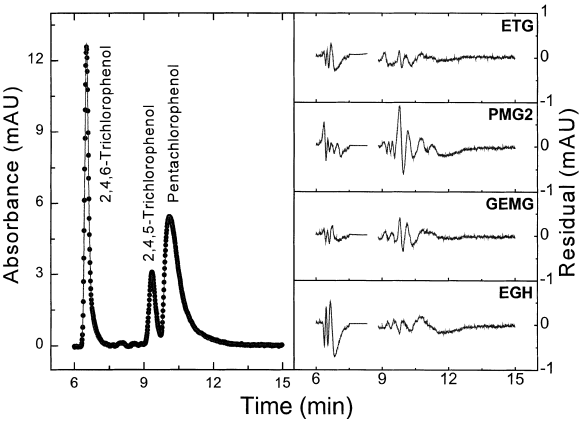

Fig. 5. Asymmetric tailing chlorophenol peaks and the fitting results. The first peak is 2,4,6-trichlorophenol, the second peak is 2,4,5-trichlorophenol, and the third peak is pentachlorophenol. The last two peaks are fitted together. All other descriptions as in Fig. 1.

ethylpyridine, and 2,6-dimethylaniline. The range of peak asymmetry is from 1.5 to 7.7. It can be seen from Figs. 3–7 and Table 3 that the ranges of the SSR ratios are from 1 to 1202 with an average of 196 for PMG2; from 0.3 to 580 with an average of 76 for GEMG; and from 2 to 278 with an average of 66 for EGH. Except for the fit of 2,4,6-trichlorophenol by GEMG peak, the ETG function offers the best for all asymmetric peaks with a wide range of asymmetry. Based on the average SSR ratios, PMG2 function seems to be the poorest function.

Although it is impossible to draw a general

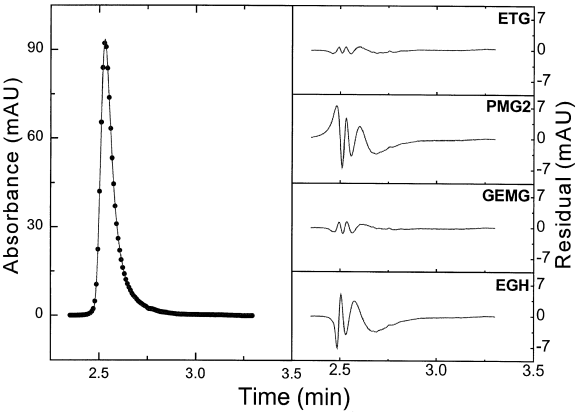

Fig. 6. Asymmetric tailing 4-ethylpyridiene peak and the fitting results. All descriptions as in Fig. 1.

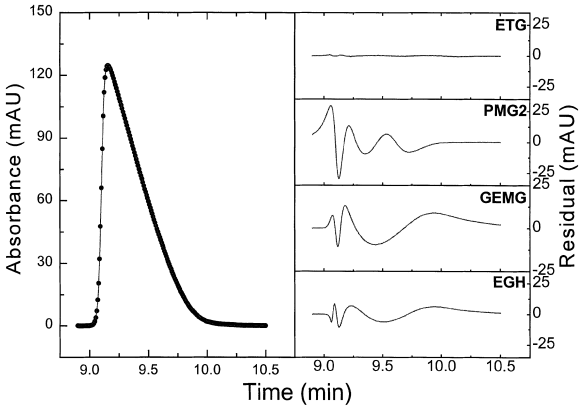

Fig. 7. Asymmetric tailing 2,6-dimethylaniline peak and the fitting results. All descriptions as in Fig. 1.

conclusion now, it appears that, by comparing the average SSR values in Table 3, the capability of the four peak functions can be arranged as: ETG> EGH>GEMG>PMG2. The relative capability of these functions is not due to the number of parameters, but their design.

5. Conclusions

Four major peak functions (ETG, PMG2, GEMG and EGH) are compared for their capability in describing real chromatographic peaks with an emphasis on the ETG function. Real chromatographic peaks of different shapes (fronting, symmetric, and tailing) are used. It is concluded that ETG function offers the highest flexibility to all shapes of chromatographic peaks. The capability of this function should allow us to better process chromatographic data. This includes the accurate determination of peak area, deconvolution of overlapped peaks, smoothing noisy peaks for the determination of statistical moments, etc.

Acknowledgements

The author would like to acknowledge the suggestion for this work from Professor Peter W. Carr at the University of Minnesota.

References

[1] J.Å. Jönsson, in: J.Å. Jönsson (Ed.), Chromatographic Theory and Basic Principle, Marcel Dekker, New York, 1987, p. 27, Chapter 2.
[2] S.N. Chesler, S.P. Cram, Anal. Chem. 43 (1971) 1922.
[3] E. Grushka, M.N. Myers, P.D. Schettler, J.C. Giddings, Anal. Chem. 41 (1969) 889.
[4] E. Grushka, in: N. Catsimpoolas (Ed.), Methods of Protein Separations, Vol. 1, Plenum Press, New York, 1975, p. 161.
[5] K. Miyabe, S. Sotoura, G.J. Guiochon, J. Chromatogr. 919 (2001) 231.
[6] K. Miyabe, G. Guiochon, Anal. Chem. 71 (1999) 889.
[7] K. Miyabe, G. Guiochon, in: P.R. Brown, E. Grushka (Eds.), Advances in Chromatography, Vol. 40, Marcel Dekker, New York, 2000, p. 1.
[8] K. Miyabe, G. Guiochon, J. Phys. Chem. B 103 (1999) 11086.
[9] M. Kubin, J. Appl. Polym. Sci. 30 (1985) 2237.
[10] J.V. Dawkins, G. Yeadom, J. Chromatogr. 188 (1980) 333.
[11] M.S. Jeansonne, J.P. Foley, J. Chromatogr. Sci. 29 (1991) 258.
[12] J. Li, Anal. Chem. 69 (1997) 4452.
[13] J.R. Torres-Lapasio, J.J. Baeza-Baeza, M.C. Garcia-Alvarez-Coque, Anal. Chem. 69 (1997) 3822.
[14] P. Nikitas, A. Pappa-Louisi, A. Papageorgiou, J. Chromatogr. 912 (2001) 13.
[15] K. Lan, J.W. Jorgenson, J. Chromatogr. 915 (2001) 1.
[16] J. Li, P.W. Carr, Anal. Chem. 69 (1997) 837.
[17] J. Li, P.W. Carr, Anal. Chem. 69 (1997) 2202.