Dependence of Chromatogram Peak Areas Obtained by Curve-Fitting on the Choice of Peak Shape Function

Mindy L. Phillips and Robert L. White*

Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019

Abstract

Seven different chromatographic peak shape functions are evaluated for use with nonlinear least-squares curve-fitting for asymmetric peak area measurements. The four-parameter functions that are evaluated are unable to adapt to peaks exhibiting substantial tailing, which results in peak area underestimates. Peak tailing is more readily represented by using six- and eight-parameter functions; however, the use of these more flexible functions can result in peak area overestimation and inaccuracies for overlapping peaks. The exponentially modified Gaussian function is advocated for general purpose chromatogram curve fitting because it adapts to extended peak tailing better than the other four-parameter functions evaluated and provides better reproducibility when large digitization intervals are employed or when poor initial curve-fitting parameters are chosen than the six- and eight-parameter functions that were evaluated.

Introduction

When quantitative analysis involves the detection of species separated by chromatography, the accuracy of the method is dictated by the accuracy with which chromatogram peak areas can be measured. The availability of powerful chromatography data stations makes it possible to calculate chromatogram peak areas in reasonably short times by using computationally intensive curve-fitting techniques. Curve-fitting algorithms match peak shape functions to digitized chromatograms by systematically adjusting peak shape function parameters. Curve-fitting methods can provide more accurate chromatogram peak area measurements than numerical integration, particularly when chromatograms contain unresolved elutions (1,2). The best results are obtained when curve-fitting functions can be adjusted to conform to a wide variety of asymmetric elution profiles (3–7).

Many different peak shape functions have been proposed for use in chromatogram curve-fitting (8,9). Some functions were derived from elution theories by considering nonideal solute elution effects (10,11). Other functions were chosen because they could be varied from a symmetric form (e.g., Gaussian distribution) to an asymmetric form by an appropriate choice of parameters (12–21). Although numerous peak shape functions are available for chromatogram curve-fitting, little has been done to evaluate how well these functions adapt to chromatographic data (4,5). Such evaluations are important because curve-fitting software provides many options for the choice of peak shape function but little guidance as to which functions are best in specific situations. Evaluations based on matching peak shape functions to actual chromatograms are of little use because true peak shape parameters are not accurately known.

The approach taken here was to compare the abilities of peak shape functions to adapt to well-defined synthetic elution curves. The synthetic chromatograms employed for comparisons were calculated by using peak shape functions that were proposed as models for chromatographic elution profiles. Seven different peak shape functions were evaluated for use in chromatogram peak area determinations by using nonlinear least-squares curve-fitting. The abilities of these functions to adapt to asymmetric peak shapes and the effects of detector noise, overlapping elutions, and chromatogram digitization interval on the accuracies of peak areas derived from curve fitting were compared.

Experimental

Programs were written in the language C and executed on IBM-compatible personal computers. The curve-fitting algorithm was based on the Levenberg–Marquardt method (22) and provided a graphic display of the fitting progress from successive iterations. The curve-fitting program was written so that the peak shape function used for fitting was defined within a subroutine which was modified to change fitting functions. A subroutine that generated random numbers was used to add noise to synthetic elution profiles. To determine the effect of

^{*} Author to whom correspondence should be addressed.

detector noise on fitting results, curvefitting was applied to five different noisy synthetic peaks generated from the same peak shape parameters. The average peak area errors and standard deviations derived from these five measurements were reported. Unless otherwise stated, synthetic chromatograms were digitized at 0.167σ intervals, where σ represents the standard deviation of the totally symmetric elution profile (i.e., no tailing).

To evaluate the degree to which different peak shape functions could adapt to asymmetric elution profiles, it was necessary to ensure that peak area discrepancies that were revealed after curve-fitting were representative of inherent differences in the fitting functions and were not due to the least-squares curve-fitting algorithm stopping at local minima. Because curve-fitting results can be heavily dependent on the choice of initial parameters, care was taken to select initial parameters that were close to the expected elution curve values. When fitting noise-free elution profiles, local leastsquares minima were avoided by periodically restarting the fitting process after changing the function parameters slightly and verifying that the parameters tended toward the same optimum values. Function parameters obtained by this procedure were then used as initial parameters when elution profiles containing 5% random noise were fit.

Results and Discussion

Mathematical expressions that were used to represent elution profiles (peak shapes) and to calculate peak areas for functions that were evaluated in this study are listed in Table I. The *c*, *t*, *A*, σ , and *t*_R symbols in Table I denote eluent concentration, time, peak amplitude, standard deviation (for the totally symmetric function), and retention time, respectively. Peak shape parameters that determine the degree of peak tailing are designated by α 's. L and R subscripts are used in Table I to distinguish parameters used for the left (front) and right (tail) sides of peaks. For convenience, peak shape functions are referred to here by the two letter designations given in Table I. Note that the BG, EM, FS, LN, and HV functions all have four independent variables and that the CG and CC functions require six and eight

Table I. Peak Shape Functions	
Bi–Gaussian (15, 16) [BG]	
$c = A \exp\left(\frac{-(t - t_R)^2}{2\sigma_L^2}\right) t < t_R$	
$c = A \exp\left(\frac{-(t - t_R)^2}{2\sigma_R^2}\right) t > t_R$	
$c = A t = t_R$	
area = $\sqrt{\frac{\pi}{2}}A(\sigma_L + \sigma_R)$	
Exponentially modified Gaussian (20) [EM]	
$c = \frac{\sqrt{2\pi}A\sigma}{2\alpha_1} \exp\left[\frac{\sigma^2}{2\alpha_1^2} + \frac{t_R - t}{\alpha_1}\right] \left[1 + erf\left(\frac{t - t_R}{\sqrt{2\sigma}} - \frac{\alpha}{\sqrt{2\alpha_1}}\right)\right]$	
area = $\sqrt{2\pi}A\alpha$	
Fraser–Suzuki (14) [FS]	
$c = A \exp\left(-\ln 2\left\{\ln \left[1 + \alpha_1 \frac{\left(t - t_R\right)}{\sigma}\right] / \alpha_1\right\}^2\right)$	
area = $\sqrt{\frac{\pi}{\ln 2}} A \sigma \exp\left[\frac{\alpha_1^2}{(4\ln 2)}\right]$	
Log-normal (21) [LN]	
$c = A \exp\left[\frac{-\ln 2}{\ln^2(\alpha_1/\sigma)}\ln^2\left(\frac{t-t_R}{\sigma+\alpha_1} \frac{(\alpha_1/\sigma)^2-1}{(\alpha_1/\sigma)}+1\right)\right]$	

area =
$$\sqrt{\frac{\pi}{\ln 2}}A(\sigma + \alpha_1) \left[\frac{\alpha_1/\sigma}{(\alpha_1/\sigma)^2 - 1}\right] \exp\left[\frac{\ln^2(\alpha_1/\sigma)}{(4\ln 2)}\right] \ln(\alpha_1/\sigma)$$

parameters, respectively. With the exception of the CC function, peak shape functions could be integrated to obtain closed form expressions for calculating areas. For the CC function, areas were calculated by numerical integration of a digitized representation of the noise-free peak shape.

To evaluate how effectively peak shape functions could adapt to different asymmetric elution profiles, synthetic peaks were generated for each of the functions listed in Table I, and these synthetic peaks were employed to test the accuracy of peak areas measured by curve-fitting. The synthetic elution profiles used in this study are shown in Figure 1. Function parameters employed to generate these unit area peaks were selected so that eluent concentration at 4σ past the peak maximum on the tailing side of the peak was about 5% of the peak maximum. As shown in Figure 1, this resulted in a range of asymmetric peak shapes with different tailing characteristics. Due to the nature of the CG and CC functions, peak tailing for these curves extended farther than for the other functions.

Peak area errors obtained by fitting each of the peak shape



functions listed in Table I to the Figure 1 elution profiles are plotted as a function of tailing peak type in Figure 2. None of the functions were able to conform to all seven synthetic elution profiles. Figure 2 shows the extent to which each function could conform to the other elution profiles. Peak area errors resulting from fitting to the BG tailing peak profile were the smallest of the seven elution curves, which indicated that the tailing BG peak shape was more readily approximated by the peak shape functions than the other tailing peak types. The largest negative peak area errors (i.e., peak areas were underestimated) resulted when the CG and CC tailing peaks were fit with the BG, EM, FS, LN, and HV peak shapes. These fourparameter peak shape functions could not accommodate the extensive peak tailing that was characteristic of the CG and CC tailing peak shapes. However, the EM function yielded significantly better fits to the CG and CC peaks than the other fourparameter functions. In general, peak area errors tended to be negative when the BG, FS, LN, and HV peak shapes were used for fitting. In contrast to those four-parameter functions, fitting

the EM function to the tailing elution curves resulted in negative area errors for the CG and CC profiles but positive area errors for the other tailing peak shapes. Peak area errors were always positive when the CG and CC functions were used for fitting the other peak shapes. These two functions consistently resulted in overestimated peak areas.

The effects of detector noise on peak area measurement accuracies derived from curve-fitting were investigated by using the functions listed in Table I to fit noisy elution curves. A series of tailing elution curves were generated from the same peak shape functions used to make the elution profiles shown in Figure 1; 5% (relative to the peak maximum) peak-to-peak random noise was added to these curves. Peak area errors obtained by fitting these noisy elution profiles are shown in Figure 3. Error bars in Figure 3 represent area calculation standard deviations obtained from the results of fitting five different noisy curves for each tailing peak type. Comparing Figures 2 and 3 reveals that peak area errors were significantly greater when noise was present. However, the general trends observed in Figure 3 are also apparent in Figure 2. Like the results obtained for the noise-free elution curves, the six-parameter CG and eight-parameter CC functions resulted in overestimated peak areas, whereas the four-parameter functions tended to result in underestimated ones. Comparing Figures 2 and 3 also reveals that fits employing the FS and LN peak shape functions yielded nearly identical results, which suggests that these two functions have similar abilities to adapt to asymmetric peak shapes.

The effect of unresolved elutions on peak area measurement accuracy was evaluated by fitting the peak shape functions to elution curves consisting of two equal-amplitude overlapping Gaussian profiles. Gaussian peak shapes were selected for overlapping elution profiles because all of the peak shape functions evaluated here could be converted to symmetric Gaussians through the proper choice of peak shape parameters. Elution profiles were generated to represent well-resolved (R = 1.0) and poorly resolved (R = 0.5) overlapping elutions. Before random noise was added to the elution curves, the errors for the individual peak areas derived from curve-fitting were less than 0.06% for all peak shape functions, which suggested that all of the functions could accurately represent Gaussian distributions. Peak area errors obtained from fitting overlapping elution profiles containing 5% random noise are listed in Table II. The peak shape functions were able to fit noisy unit resolution elution curves well but, with the exception of the CG fitting function, exhibited greater variability (higher



standard deviations) for the first peak than for the second peak. Because peak tailing was not incorporated into synthetic curves, peak area calculation errors and standard deviations for both peaks would be expected to be similar. The larger peak area calculation standard deviation obtained for the first peak in the R = 1.0 curve may have been an artifact associated with the curve-fitting algorithm. However, the trend in peak area calculation standard deviation observed for the R = 1.0 fitting results was not apparent in the R = 0.5 results. Results given in Table II also show that the FS function failed to accurately represent the constituent peaks in the R = 0.5 elution curve. When the FS function was employed for curve-fitting, the area of one peak was consistently underestimated and the area of the other peak was consistently overestimated. However, the magnitude of the underestimation was about the same as that of the overestimation, so the total curve area was significantly more accurate than the individual peak areas. The CG fitting function produced the worst results for the R = 0.5 elution curves. Unlike the results obtained by using the FS fitting function, average peak area errors were positive for both over-

> lapping peaks when the CG function was employed because areas of both peaks were typically overestimated.

> The digitization interval used to generate chromatograms can also influence curvefitting accuracy. When flame ionization or thermal conductivity detectors are employed in gas chromatography or absorbance detection is employed in liquid chromatography, short digitization intervals are commonly used to accurately represent elution profiles by digitized detector signals. However, much longer digitization intervals are often used when mass spectrometers are used as chromatography detectors because this permits greater mass spectral signal averaging for each acquired spectrum. To evaluate the effect of digitization interval on the accuracy and precision of peak area measurements obtained by curvefitting, curve-fitting was applied to noisy synthetic elution curves that were digitized at 1σ and 2σ intervals. When digitized at 1σ intervals, peaks were defined by 6–7 points, whereas 3–4 points defined peak shapes when curves were digitized at 2σ intervals. The elution curves used for these evaluations were generated by using the same functions and parameters employed to make the curves shown in Figure 1, and 5% random noise was added to each synthetic chromatogram. Peak area errors obtained by curve-fitting elution curves digitized at 1σ intervals are plotted as a function of tailing peak type in Figure 4. Trends similar to those found in Figure 3 are also present in Figure 4, but the peak area error standard deviations are significantly higher

in Figure 4 than in Figure 3. In particular, Figure 4 shows that the EM fitting function still yielded better fits to extensively tailing peak shapes (CG and CC tailing peak types) than the other four-parameter fitting functions when the digitization interval was increased from 0.167σ (Figure 3) to 1σ (Figure 4). Figure 5 contains plots of peak area errors as a function of tailing peak type derived from curve-fitting elution curves that were digitized at 2σ intervals. The general trends found in Figures 2–4 are not apparent in Figure 5, and the peak area errors and peak area error standard

deviations are much larger in Figure 5 than in Figure 4. Interestingly, peak area errors and peak area error standard deviations obtained by fitting BG tailing elution curves were much smaller than those obtained by fitting the other elution curve types.

Peak area calculation standard deviations derived from elution curves digitized at 1σ and 2σ intervals, which represent measurement reproducibilities, are plotted as a function of tailing peak type in Figures 6 and 7, respectively. When the elution profile was based on a tailing BG peak shape, little change in peak area measurement precision was observed after doubling the digitization interval. More dramatic changes were observed for fits of elution curves derived from the other peak shapes. In general, the precision of peak area measurements decreased (standard deviation increased) when the digitiza-



Figure 2. Peak area calculation errors derived from fitting each peak shape function to seven asymmetric elution curves. Results from individual fitting functions are represented as follows: \bullet = BG, \bigcirc = EM, \blacksquare = FS, \square = LN, \blacktriangle = HV, \blacktriangledown = CG, and \diamondsuit = CC.

Table II. Average Peak Area Errors and Standard Deviations for OverlappingElutions Containing 5% Noise

	0				
	R = 0.5		R =	<i>R</i> = 1.0	
Function	Peak 1 (%)	Peak 2 (%)	Peak 1 (%)	Peak 2 (%)	
BG	-0.15 ± 0.90	0.19 ± 0.89	-0.12 ± 0.76	0.00 ± 0.46	
EM	0.00 ± 0.49	0.05 ± 0.88	-0.20 ± 0.94	-0.03 ± 0.43	
FS	-0.76 ± 1.15	0.85 ± 0.93	-0.05 ± 0.67	-0.11 ± 0.42	
LN	0.10 ± 0.52	-0.07 ± 0.85	0.00 ± 0.78	-0.05 ± 0.33	
HV	-0.03 ± 0.75	0.14 ± 0.68	-0.26 ± 0.89	0.00 ± 0.38	
CG	1.11 ± 1.54	0.55 ± 0.45	0.01 ± 0.70	0.54 ± 1.01	
CC	0.18 ± 0.67	-0.20 ± 0.98	-0.25 ± 0.82	-0.10 ± 0.33	

tion interval was increased. However, the most significant increases in peak area measurement standard deviation were associated with the CG and BG fitting functions, particularly when the FS and LN elution profiles were fit.

Conclusion

Effects of the choice of peak shape function on chromatogram curve-fitting results are difficult to predict. However, it is clear that peak area measurement accuracy depends on the choice of peak shape function and the form of the chromatographic elution profile. The comparisons of curve-fitting



Figure 3. Peak area calculation errors derived from fitting each peak shape function to seven asymmetric elution curves to which 5% peak-to-peak random noise had been added. Results from individual fitting functions are represented as follows: ● = BG, O = EM, ■ = FS, □ = LN, ▲ = HV, ▼ = CG, and \diamondsuit = CC.





results reported here do not represent a comprehensive study of the effect of peak shape function on curve-fitting efficiency. However, some general trends were revealed by this study. All of the four-parameter functions evaluated in this study were incapable of representing asymmetric peaks with extensive tailing. Use of the six-parameter CG function always resulted in peak area overestimates. This peak shape function was the most sensitive to chromatogram noise and digitization interval, and the precision with which this function could be used to fit overlapping peaks was lower than for the other evaluated functions. Based on these findings, the CC function might be considered to be the best peak shape choice for fitting chromatograms of those evaluated here. However, because the CC function had more degrees of freedom than any of the other evaluated functions, it was more susceptible to curve-fitting anomalies due to inappropriate choices for initial function parameters. For example, when the CC function was used to fit an elution curve that consisted of two noise-free Gaussian peaks (R = 0.5) and initial parameters that were substantially different from the correct values were chosen, the best fit resulted in a total area error that was 0.2%, but the individual peak areas errors were 16% and -16%, respectively. Errors this large were not obtained when poor initial parameter selections were used with the less flexible four-parameter functions. Of the four-parameter functions evaluated, the FS and LN provided the best fits to the other four-parameter tailing peaks, but the EM function provided the best fits to the CG and CC tailing peak shapes. Based on the comparisons described here and the fact that fits employing the CC function were



Figure 6. Peak area calculation standard deviations resulting from fitting elution profiles digitized at 1σ intervals. Results from individual fitting functions are represented as follows: $\Phi = BG$, O = EM, $\blacksquare = FS$, $\square = LN$, $\blacktriangle = HV$, $\blacktriangledown = CG$, and $\diamondsuit = CC$.



more dependent on the choice of initial parameters than when the four-parameter functions were employed, the exponentially modified Gaussian (EM) function provided the best overall performance (of those evaluated in this study) for quantitative analysis when chromatographic curve-fitting was employed.

References

- 1. A.W. Westerberg. Detection and resolution of overlapped peaks for an on-line computer system for gas chromatographs. *Anal. Chem.* **40**: 1770–77 (1969).
- V.R. Meyer. Errors in the area determination of incompletely resolved chromatographic peaks. J. Chromatogr. Sci. 33: 26–33 (1995).
- J. Grimalt, H. Iturriaga, and X. Tomas. The resolution of chromatograms with overlapping peaks by means of different statistical functions. *Anal. Chim. Acta* **139**: 155–66 (1982).

- 4. J. Grimalt, H. Iturriaga, and J. Olivé. An experimental study of the efficiency of different statistical functions for the resolution of chromatograms with overlapping peaks. *Anal. Chim. Acta* **201**: 193–205 (1987).
- J. Olivé, J.O. Grimalt, and H. Iturriaga. Resolution of overlapping peaks in gas and liquid chromatography, evaluation of different statistical and empirical functions. *Anal. Chim. Acta* **219**: 257–72 (1989).
- D.Y. Youn, S.J. Yun, and K.H. Jung. Improved algorithm for resolution of overlapped asymmetric chromatographic peaks. *J. Chro*matogr. 591: 19–29 (1992).
- J. Olivé and J.O. Grimalt. Relationships between different chromatographic peak description functions and numerical solutions of the mass balance equation. J. Chromatogr. Sci. 33: 194–203 (1995).
- E. Reh. An algorithm for peak-shape analysis for differentiating unresolved peaks in chromatography. *Trends Anal. Chem.* 12: 192–94 (1993).
- E. Reh. Peak-shape analysis for unresolved peaks in chromatography: Comparison of algorithms. *Trends Anal. Chem.* 14: 1–5 (1995).
- 10. J.C. Giddings. Kinetic origin of tailing in chromatography. *Anal. Chem.* **35**: 1999–2002 (1963).
- P.C. Haarhoff and H.J. van der Linde. Concentration dependence of elution curves in nonideal gas chromatography. *Anal. Chem.* 38: 573–81 (1966).
- J. Pitha and R.N. Jones. A comparison of optimization methods for fitting curves to infrared band envelopes. *Can. J. Chem.* 44: 3031–50 (1966).
- 13. J. Pitha and R.N. Jones. An evaluation of mathematical functions to fit infrared band envelopes. *Can. J. Chem.* **45**: 2347–52 (1967).
- R.D.B. Fraser and E. Suzuki. Resolution of overlapping bands: Functions for simulating band shapes. *Anal. Chem.* 41: 37–39 (1969).
- 15. E. Grushka, M.N. Myers, and J.C. Giddings. Moments analysis for the discernment of overlapping chromatographic peaks. *Anal. Chem.* **42:** 21–26 (1970).
- T.S. Buys and K. de Clerk. Bi-Gaussian fitting of skewed peaks. Anal. Chem. 44: 1273–75 (1972).
- S.N. Chesler and S.P. Cram. Effect of peak sensing and random noise on the precision and accuracy of statistical moment analyses from digital chromatographic data. *Anal. Chem.* 43: 1922–33 (1971).
- S.N. Chesler and S.P. Cram. Iterative curve fitting of chromatographic peaks. *Anal. Chem.* 45: 1354–59 (1973).
- F. Dondi, A. Betti, G. Blo, and C. Bighi. Statistical analysis of gas chromatographic peaks by the Gram-Charlier series of type A and the Edgeworth-Cramer series. *Anal. Chem.* 53: 496–504 (1981).
- J.P. Foley and J.G. Dorsey. A review of the exponentially modified Gaussian (EMG) function: Evaluation and subsequent calculation of universal data. J. Chromatogr. Sci. 22: 40–46 (1984).
- J.O. Grimalt and J. Olivé. Log-normal derived equations for the determination of chromatographic peak parameters from graphical measurements. *Anal. Chim. Acta* 248: 59–70 (1991).
- D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11: 431–41 (1963).

Manuscript accepted September 16, 1996.