Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis
33 (2003) 563–570
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

# Use of mathematical derivatives (time-domain differentiation) on chromatographic data to enhance the detection and quantification of an unknown 'rider' peak

S.J. Ford *, M.A. Elliott, G.W. Halbert

*Cancer Research UK Formulation Unit, Department of Pharmaceutical Sciences, University of Strathclyde, 204 George Street, Glasgow, Scotland G1 1XW, UK*

Received 16 January 2003; received in revised form 26 May 2003; accepted 28 May 2003

**Abstract**

Two samples of an anticancer prodrug, AQ4N, were submitted for HPLC assay and showed an unidentified impurity that eluted as a 'rider' on the tail of the main peak. Mathematical derivatization of the chromatograms offered several advantages over conventional skimmed integration. A combination of the second derivative amplitude and simple linear regression gave a novel method for estimating the true peak area of the impurity peak. All the calculation steps were carried out using a widely available spreadsheet program.


*Keywords:* Tangent integration; Skimmed integration; Resolution enhancement; Baseline prediction

## 1. Introduction

'Rider' peaks appear in a chromatogram when a small detected band is incompletely resolved from a larger peak. The term usually refers to peaks, which elute on the tail of a main peak. The accurate characterisation and quantification of rider peaks is known to be difficult and several alternatives have been investigated. Triangular/perpendicular methods and valley-to-baseline for rider integration are known to be prone to errors [1,2]. The popular tangential (skimming) method requires careful selection of the baseline end points [3], however, significant errors can still occur [4,5]. The use of mathematical deconvolution of the chromatographic data has been studied [6] as have 2D calibration techniques [7].

Chromatographic data is derivatized as part of the peak integration process [1,3] and the use of derivative chromatography for the quantification of closely eluting peaks of similar magnitude has been studied previously [8,9]. Theoretical calculations have showed good results for peak area ratios between 100 and 10% [10]. Grushka and co-workers have used the second derivatives of theoretical peaks over a wide range of relative peak areas to determine the start and stop points for skimmed integration [11,12]. Fully resolved

* Corresponding author. Fax: +44-141-548-4903.
*E-mail address:* crcfus@strath.ac.uk (S.J. Ford).

0731-7085/03/$ - see front matter 
doi:10.1016/S0731-7085(03)00354-6

second derivative curves have been shown to occur in theoretical Gaussian systems at 'rider' peak areas of 0.01% [11].

The Cancer Research UK Formulation Unit was provided with two samples of a new anticancer prodrug, AQ4N (1,4-bis[2-(dimethylamino)ethyl] amino-5,8-dihydroxyanthracene-9,10-dione-bis-*N*-oxide) [13], for the purposes of formulation and analytical development. During HPLC assay development a small unknown impurity (nominally labelled Impurity D) eluted as a rider on the tail of the main peak. The level of Impurity D varied slightly between the samples. We were not able to improve the resolution of the parent compound and Impurity D peaks: the developed HPLC assay method appeared to give the optimal separation. These investigations were hampered by the lack of an appropriate standard and the low Impurity D content in the AQ4N samples. There was no discernable difference in the UV profiles of AQ4N and Impurity D.

This study investigates the use of second derivatives of experimental data to obtain quantitative information on a 'rider' peak. The work confirms previous theoretical predictions of the advantages that second derivatives offer and the high errors associated with 'skimmed' integration. Furthermore, a novel analysis is carried out where the second derivative information is used as a 'marker' for rider concentration allowing linear regression of the chromatographic traces and prediction of an appropriate baseline. This is done by comparatively simple calculations within a spreadsheet program widely available on computer-based HPLC systems.

## 2. Experimental

### 2.1. Chemicals

HPLC grade reagents and solvents were used throughout. Two samples of AQ4N·2HCl (nominally labelled samples 'A' and 'B') were provided by BTG in collaboration with Denny (Auckland Cancer Society) [14].

### 2.2. Chromatography

The chromatographic analysis was carried out on a TSP HPLC system (ThermoFinnigan, Hemel Hempstead, UK), consisting of a SM4000 four line vacuum degasser, P2000 binary gradient pump, A1000 autosampler and a UV1000 detector integrated via a SN4000 SpectraNet module with a PC (Dell Optiplec Gm) running HPLC acquisition software PC1000 (version 3.0.1). The HPLC system is calibrated on a bi-annual basis.

Mobile phase A consisted of 0.1% (v/v) trifluoroacetic acid in water–acetonitrile (75:25, v/v). Mobile phase B was acetonitrile. The gradient used was 100% A for 10 min, with a linear gradient to 100% B at 20 min followed by a 16 min equilibration period. The column was a Phenomenex (UK) Luna C8(2) (5 μm, 150 × 4.6 mm) dedicated to the AQ4N HPLC assay. The flow rate was 1 ml/min, with an injection volume of 20 μl. The UV detection wavelength was 245 nm. Each sample was injected in triplicate.

The integration of the Impurity D peak was carried out used the 'Rider' integration option in the PC1000 software. The start and end points of the baseline were chosen manually according to the guidelines suggested by Dyson [3]. Chromatograms were exported as comma separated (CSV) files under the PC1000 Data Maintenance program using start/stop times of 4.3/5.5 min and a data interval of 1.

### 2.3. Sample preparation

Samples A and B were prepared at 0.2 mg/ml in sodium orthophosphate buffer (pH 7.0, 10 mM). Mixtures of these two samples were prepared using calibrated Gilson variable volume pipettes in the following proportions: 100A:0B, 75A:25B, 50A:50B, 25A:75B and 0A:100B.

### 2.4. Computational calculations

All the computational calculations were carried out using Microsoft EXCEL 98 (Macintosh Edition). Second derivative amplitude maxima, linear regression and calculation of S.D. were all carried

out using the appropriate functions within the spreadsheet.

### 2.5. Differentiation calculations

The derivatization of the chromatography data was carried out using a simple Savitzky–Golay series for the second derivative of a quadratic/cubic polynomial over 11 points [15]. The equation used to obtain the second derivative was as follows:

$$y'' = -\frac{1}{429\Delta t^2} \times (15y_{i-5} + 6y_{i-4} - y_{i-3} - 6y_{i-2} - 9y_{i-1} - 10y_i - 9y_{i+1} - 6y_{i+2} - y_{i+3} + 6y_{i+4} + 15y_{i+5})$$

where $y''$, second order derivative of the chromatogram at point $y_i$; $\Delta t$, time interval; and $y_{i-5}$ to $y_{i+5}$, consecutive points on the chromatogram.

This particular series of Savitzky–Golay parameters were chosen since a quadratic is the minimum order equation required to obtain a second derivative, while eleven points provided the optimum smoothing for the data points concerned. Prior to the application of the Savitzky–Golay algorithm, initial data bunching was carried out as an averaging process of six sequential datapoints, this generated a single time point every 0.005 min. This reduced the size of the subsequent spreadsheets and acted to smooth the data. A simpler, but more cumbersome, stepwise numerical derivatization process of consecutively calculating the gradient of a slope between adjacent points was also assessed; it generated equivalent results to the Savitzky–Golay process but required an additional averaging step and larger spreadsheets.

A typical second derivative curve would contain two maxima and one minimum. For the rider peaks shown here the low retention time maximum is obscured by the signal from the main peak. All the second derivative amplitude results come from the high retention time maximum of the Impurity D curve (indicated by the arrow in Fig. 2) where the interference from the main peak is minimized.

### 2.6. Theoretical calculations

The Gaussian equation used to describe model chromatographic peaks is as follows [2]:

$$h(t) = \frac{A}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(t - t_R)^2}{2\sigma^2}\right]$$

where, $h(t)$, peak height at time $t$; $A$, total peak area; $t_R$, time at peak maximum (retention time), and $\sigma$, S.D. of the peak.

The theoretical model of the peak complex was generated by the addition of two Gaussian functions. The total peak area values ($A$) were selected so that the ratio of the $A$ values reflected the approximate peak area ratio determined experimentally. The total peak area values for rider peak were varied to investigate the response of the second derivative amplitude. The $t_R$ values were set to the retention times of the main peak and Impurity D and the $\sigma$ values were optimised to generate a peak complex, which visually represented the experimental traces. The dimensionless Gaussian parameters are shown in Table 1. The theoretical data was processed in the same way as the experimental data described above without the initial bunching operation.

Skimmed integration of the theoretical peak complex was carried out in the Excel spreadsheet by selecting two 'baseline' points on the curve and fitting a straight line. Peak areas were calculated by subtracting a baseline from the curve, multiplying by time interval and summing over the relevant range.

Table 1
Gaussian parameters for the theoretical curves

| Parameter | Values for Peak 1 | Values for Peak 2 |
|---|---|---|
| $A$ | 10 000 | 10, 20, 30, 40, 50 |
| $t_R$ | 4.0 | 4.7 |
| $\sigma$ | 0.16 | 0.15 |

## 3. Results and discussion

### 3.1. Skimmed integration and second derivative results

A typical chromatogram of the AQ4N and Impurity D peak in Sample B is shown in Fig. 1. The Impurity D peak and its corresponding second derivative trace are shown in Fig. 2.

The mean and S.D. values for peak areas given by skimmed integration and second derivative amplitude of Impurity D peak against sample mixture composition are shown Fig. 3. The equivalent mean values for theoretical peak areas against the rider peak area parameter ($A$) are shown in Fig. 4. The parameters generated by linear regression of all four sets of data are shown in Table 2.

The experimental and theoretical results show similar trends. The non-linear results from the peak area data originate from the invalid use of straight 'baseline' in a situation where the peak clearly sits on top of a 'curved' background. The non-linear response for the skimmed peak area suggests that even if an Impurity D (or other rider) standard were available, a standard additions approach to quantification would generate erroneous results. The theoretical results show that the skimmed integration method severely underestimates the actual peak area as found previously [3–5].

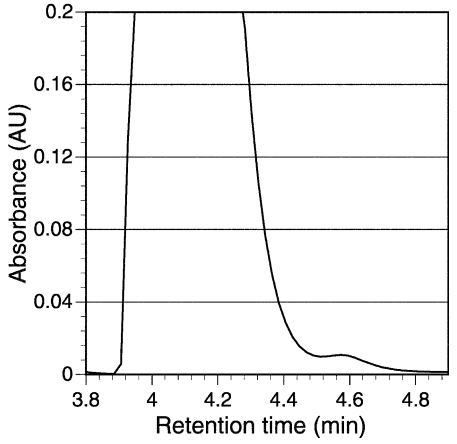


Fig. 1. Typical chromatogram of the AQ4N and Impurity D peak in Sample B.

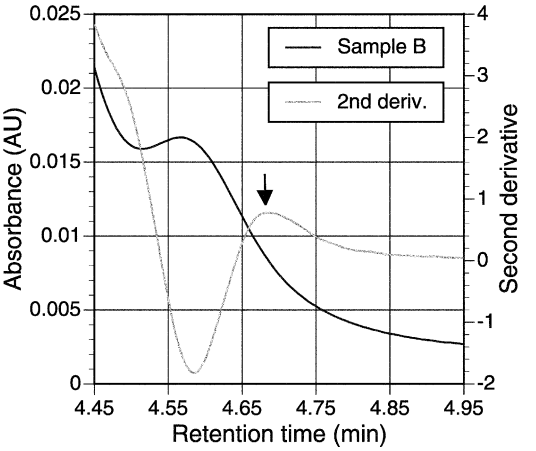


Fig. 2. Expanded chromatogram and second derivative of the Impurity D peak in Sample B. (The arrow indicates the maximum used to establish second derivative amplitude.)

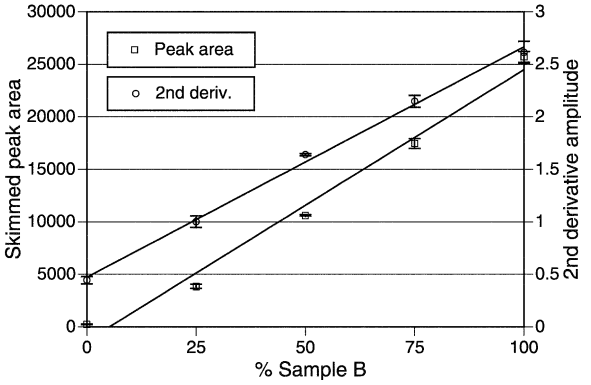


Fig. 3. Mean values (with S.D. error bars) and best fit lines for the experimentally determined values of the skimmed peak areas and second derivative amplitudes of Impurity D from Sample A:B mixtures.

The second derivative data suggests that Sample A has 17% of the Impurity D content of Sample B. The equivalent figure derived from the skimmed peak area measurements is $<1\%$: a substantial difference.

Skimmed integration is an operation which is usually difficult to reproduce: different software packages or human chromatographers may choose different start and end points on the trace. The second derivative amplitude is more specific since

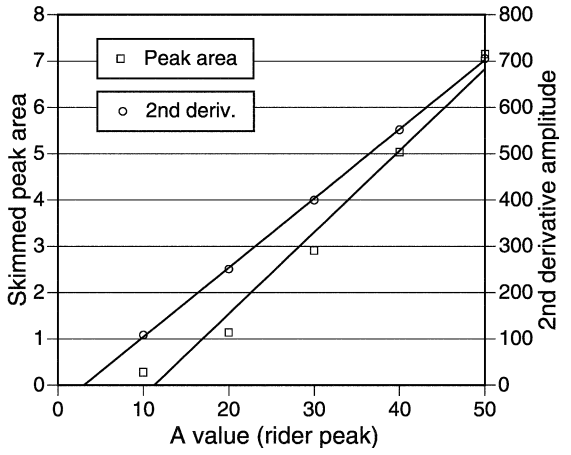


Fig. 4. Skimmed peak areas and second derivative amplitudes (with best fit lines) against theoretical peak area parameter ($A$) for the rider peak.

it is an easier parameter to define. The technique of using a second derivative curve to assist in the selection of start and end points for skimmed integration [11] would not be applicable to this data since it requires both second derivative maxima from the rider peak to be observable.

The experimental second derivative data gives a higher S.D. than the peak area data, but this is to be expected since the increased resolution of derivatized experimental data is at the expense of exaggerated experimental noise.

In order to establish a curved baseline in the absence of a drug sample completely devoid of Impurity D, the peaks would have to be fitted to known equations [6]. The use of the actual experimental data to determine the correct curved baseline by a simple linear regression is discussed below.

The good linear fit for the experimental second derivative data is mirrored by a similar result for the theoretical results, where the correlation coefficient is 1.000. This indicates that, to a reasonable approximation, the amplitude of the second derivative is proportional to Impurity D concentration. However, the negative regression intercept shows that the main peak is distorting the second derivative signal of the rider peak. This concurs with previous work [10–12].

### 3.2. Extrapolation of experimental data to obtain 'pure' drug curve

Since Samples A and B have significantly different contents of Impurity D, it is possible to extrapolate by linear regression of the available chromatograms to a point where the Impurity D content is zero, thus providing a chromatographic baseline. Two key operations are required for this process; firstly the correct alignment of the chromatograms and, secondly, assigning the Impurity D content of the sample mixtures. The latter problem is addressed by the assertion that, although the exact concentration of Impurity D cannot be deduced, the relative concentration is proportional to the second derivative amplitude.

The alignment of the chromatograms in the time axis was carried out using two 'markers'; the second derivative maximum for the Impurity D peak and the steep tail gradient of the main drug. The alignment was carried out manually within the spreadsheet. There was no restriction on the

Table 2
Parameters from the linear regression of the skimmed peak area and second derivative for experimental and theoretical results

| Regression parameter | Skimmed peak area (experimental) | Second derivative amplitude (experimental) | Skimmed peak area (theoretical) | Second derivative amplitude (theoretical) |
|---|---|---|---|---|
| Slope | 258 | 0.0219 | 0.173 | 15.0 |
| Intercept | −1337 | 0.473 | −1.85 | 45.4 |
| S.D. (slope) | 9.49 | 0.000526 | 0.0130 | 0.130 |
| S.D. (intercept) | 577 | 0.0322 | 0.431 | 4.31 |
| Standard error | 1290 | 0.072 | 0.411 | 4.10 |
| No. data points | 15 | 15 | 5 | 5 |
| $r^2$ | 0.983 | 0.993 | 0.983 | 1.000 |

displacement of the chromatograms along the time axis while alignment was being carried out using the second derivative maximum 'marker'. However, the alignment carried out using the tail gradient of the main peak was assessed visually and was considered as 'fine tuning'. Accordingly, this latter alignment operation was restricted to displacement by three data points (the equivalent of 0.015 min). After alignment, the replicate chromatographic traces from each sample were averaged.

The linear regression of the averaged chromatograms was carried out by assigning the mean second derivative amplitude for each sample mixture ($\partial_B^{2(\mathrm{AMP})}$) of the second derivative as the $x$ abscissa, and the absorbance of the traces at a fixed time point, $T$, ($A_B^T$) as the $y$ ordinate. Since five sample mixtures were used ($B = 0$, 25, 50, 75 and 100) the procedure defines five data points for each time point $T$: ($\partial_0^{2(\mathrm{AMP})}$, $A_0^T$), ($\partial_{25}^{2(\mathrm{AMP})}$, $A_{25}^T$)...($\partial_{100}^{2(\mathrm{AMP})}$, $A_{100}^T$). The resulting intercept at time $T$ where the $x$ ordinate ($\partial^{2(\mathrm{AMP})}$), and accordingly the Impurity D concentration, are zero gives the absorbance of the baseline chromatogram at time $T$. Repeating this procedure for all the relevant time points (or values of $T$) gives the baseline chromatogram required. This process predicts a curved baseline using the experimental data, without requiring peak deconvolution.

The results for the regression are shown in Fig. 5 together with the correlation coefficients of the analyses at the different time points. The correlation coefficient values show that the linear fit for the Impurity D peak to the second derivative amplitude is good. On either side of the Impurity D peak the correlation coefficients value drops, since the relationship between the chromatograms is random. The peak areas of Impurity D as given by the skimmed integration process and above extrapolated baseline are shown in Table 3. The equivalent results for the theoretical peak complex are shown in Table 4. The comparison of the skimmed integration with the extrapolated baseline figures shows that the latter approach is several times more accurate, due to the use of a curved baseline. Although the extrapolation method requires several samples, these can be

Table 3
Peak area from baseline extrapolation and skimmed integration algorithms

| Sample mixture | Peak area from baseline extrapolation | Peak area from skim integration |
|---|---|---|
| 100B:0A | 77 433 | 25 712 |
| 75B:25A | 61 236 | 17 463 |
| 50B:50A | 48 462 | 10 613 |
| 25B:75A | 30 122 | 3845 |
| 0B:100A | 12 880 | 242 |

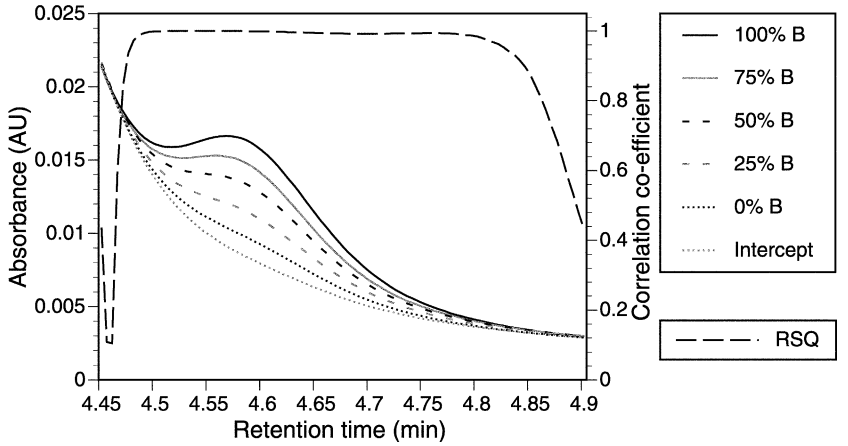


Fig. 5. Aligned chromatograms, baseline chromatogram (intercept) and correlation coefficient (RSQ) for the extrapolation (linear) regression analysis.

Table 4
Peak area from curve extrapolation on theoretical data

| $A$ value (rider peak area) | Peak area from skim integration | Peak area from curve subtraction |
|---|---|---|
| 50 | 7.15 (14%) | 46.8 (94%) |
| 40 | 5.03 (13%) | 36.8 (92%) |
| 30 | 2.91 (10%) | 26.9 (90%) |
| 20 | 1.41 (7.1%) | 16.9 (84%) |
| 10 | 0.28 (2.8%) | 6.9 (69%) |

The figures in parenthesis show the percentage of the theoretical peak area parameter, $A$, for the rider peak.

prepared by combination as long as the original samples contain significantly different rider concentration.

The errors shown in the theoretical results originate from the use of the second derivative amplitude as the $x$ ordinate in the regression, which, as highlighted above, has only an approximately linear response to rider concentration. An equivalent extrapolation analysis was also carried out using rider peak area as the $x$ abscissa (data not shown). In this analysis the theoretical peak areas could all be recovered to within 99.5% of their original values. This indicates that the errors for the baseline extrapolation shown in parenthesis in Table 4 originate from the use of the second derivative amplitude as the $x$ ordinate in the regression.

### 3.3. Peak area prediction using only derivative chromatograms

Additional studies were carried out to examine the possibility of reconstructing the Impurity D peak using a single Gaussian function and its associated second order derivative functions. The $\sigma$ value of a Gaussian function can be derived readily from its second order curve. The maximum amplitude of the second order curve is a function of the $\sigma$ value and the overall Gaussian peak area, hence the latter can be predicted. For the theoretical peak complex models described above, the correct peak area could be calculated for all five rider peak areas to within 99.5% of the actual value. However, for the experimental data the predicted values of $\sigma$ from replicate sample injections varied widely due to experimental noise. In the second derivative Gaussian function the actual peak area is proportional to $\sigma^3$ and so any errors in the latter are magnified during the subsequent mathematical processing. Although theoretically a powerful technique, this method did not assist in predicting experimental peak area.

## 4. Conclusion

The use of second derivative amplitudes has greatly assisted in the characterisation of an unknown impurity (Impurity D) which elutes as a 'rider' on the tail of AQ4N during an HPLC analysis. An estimate of peak area was obtained by using a simple linear regression procedure. The results have shown the futility of using skimmed integration in this situation and the accuracy of the second derivative approach. All the calculations were carried out in a widely used spreadsheet program and avoided the use of complex algorithms and statistical packages.

Although the use of second derivative chromatographic traces is not a replacement for the development of an HPLC method with sufficient resolving power it does provide a useful alternative for situations where 'rider' peaks occur.

## Acknowledgements

This work in this laboratory is funded by Cancer Research UK.

## References

[1] A.W. Westerberg, Anal. Chem. 41 (1969) 1770–1777.
[2] N. Dyson, J. Chromatogr. A 842 (1999) 321–340.
[3] N. Dyson, Chromatographic Integration Methods, Royal Society of Chemistry Monographs, Cambridge, 1990.
[4] V.R. Meyer, J. Chromatogr. Sci. 33 (1995) 26–33.
[5] V.R. Meyer, Chromtagraphia 40 (1995) 15–22.
[6] M. Johansson, M. Berglund, D.C. Baxter, Spectrochim. Acta 48B (1993) 1393–1409.
[7] S. Jurt, M. Schar, V.R. Meyer, J. Chromatogr. A 929 (2001) 165–168.

[8] A.A. Fasanmade, A.F. Fell, Anal. Chem. 61 (1989) 720–728.
[9] A.M. Delapena, F. Salinas, T. Galeano, A. Guiberteau, Anal. Chim. Acta 234 (1990) 263–267.
[10] E. Grushka, G.C. Monacelli, Anal. Chem. 44 (1972) 484–489.
[11] E. Grushka, D. Israeli, Anal. Chem. 62 (1990) 717–721.
[12] E. Grushka, I. Atamna, Chromtagraphia 24 (1987) 226–232.
[13] L.H. Patterson, S.R. McKeown, Br. J. Cancer 83 (2000) 1589–1593.
[14] H.H. Lee, W.A. Denny, J. Chem. Soc. Perkin Trans. 1 (1999) 2755–2758.
[15] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.