JOURNAL OF MASS SPECTROMETRY, VOL. 32, 1212–1218 (1997)

Peak Measurement in Gas Chromatographic/Mass Spectrometric Isotope Studies

L. J. C. Bluck* and W. A. Coward
Dunn Nutrition Unit, Medical Research Council, Downhams Lane, Milton Road, Cambridge CB4 1XJ, UK

A novel algorithm is described for the improved estimation of isotopomer ratios from gas chromatographic/mass spectrometric data. The minor isotopomer peak profile is fitted to an admixture of the major profile and its first and second derivatives. By this means allowance is made for non-coincidence in the measurement of the two profiles. It is shown that a simple approximation for the derivative functions allows linear methods to be used to fit the digitized data for the minor to the data for the major and to that data shifted by one unit in both directions along the x-axis. By including the x-axis vector itself in the fit data superimposed on a linear baseline (i.e. with both slope and offset) can be accommodated. The algorithm has been tested with both simulated data and real data obtained from a commercially available instrument. In addition, the data were analysed by traditional summation estimates of spectral intensity and by fitting to a particular lineshape, the exponentially modified Gaussian. In all cases the proposed method is as good as, if not better than, the others, which, in conjunction with its simplicity, makes it a useful tool for the estimation of isotopomer ratios, with potential to improve precision of tracer techniques in metabolic studies. © **1997 John Wiley & Sons, Ltd.**

J. Mass Spectrom. **32**, 1212–1218 (1997)
No. of Figures: 6 No. of Tables: 3 No. of References: 8

KEYWORDS: gas chromatography/mass spectrometry; isotopomer; stable isotopes; curve fitting

INTRODUCTION

The combination of stable isotope labelling and gas chromatographic/mass spectrometric (GC/MS) detection offers a potentially powerful method of analysis for in vivo tracer studies. The isotopomer abundances are calculated from measurements made by monitoring characteristic ions, such as a molecular grouping M, M + 1, M + 2, using the mass spectrometer in the selected ion monitoring mode. The area under the peak is then usually taken as a measure of spectral intensity but there is no absolute scaling available and it is convention to express the intensities as fractions of the intensity of one particular ion. This leads directly to the concept of isotopomer ratios, which, although they may not be the best measure of isotope abundance,[1] are universally used as the starting point for the determination of isotopic composition.

In order to determine the area under any peak, decisions must be made first about where the peak starts and ends and second about the contributions from backgrounds on which the peak is superimposed. Incorrect assessments of these parameters can lead to significant errors in determining the peak area and in other branches of spectroscopy considerable attention has been paid to these details. In GC/MS, the elimination of such errors is essential when dealing with relatively small amounts of enrichment, as often happens in human studies because the dose of isotope given is restricted on either ethical or economic grounds. It follows, therefore, that improvements in isotopomer ratio measurement will broaden the usefulness of GC/MS in tracer studies.

Commercial GC/MS instruments are normally supplied with integration algorithms, but our experience is that they are not always fully or even partially documented. However, they invariably require the operator to make some assessment of the spectrum by way of selecting values for variables to assist with the integration process; there is therefore a degree of subjectivity involved in the calculations. In addition, the estimation of peak onset and of background effects requires that there is no spectral overlap and that the peaks in question be well defined; under experimental conditions this is not always the case. For example, the chromatograms may be crowded and the operator cannot then observe a sufficiently large region of pure baseline to make a satisfactory choice.

To circumvent many of these problems, we have developed an alternative method of assessing isotopomer abundances from recorded spectra. It compensates for backgrounds, does not require that the wings of the peak be observed and no potentially subjective operator input is required. In the present work the method has been implemented on data produced by instruments from two different manufacturers and the results were compared with those obtained with the standard instrument algorithms.

* Correspondence to: L. J. C. Bluck, Dunn Nutrition Unit, Medical Research Council, Downhams Lane, Milton Road, Cambridge CB4 1XJ, UK.

CCC 1076–5174/97/111212–07 $17.50

Received May 23, 1997
Accepted July 28, 1997

METHODS

Mathematical concepts

The various methods which have been proposed for the quantitation of chromatograph output may be classed as either summation methods or fitting methods. An example of a summation method is the set of algorithms supplied with the Finnigan MAT INCOS™ system and described in the accompanying manual.[3] The level of noise associated with a chromatogram is assessed by consideration of its second derivative. The digitized chromatogram is then divided into two sets of data points: those deemed to be baseline points and those which are not baseline points. A baseline point is defined as one for which there is no straight line drawn between any point in its left-hand neighbourhood and any point in its right-hand neighbourhood which would pass underneath the point itself. This procedure is subjective because the size of the neighbourhoods is selected by the operator. The baseline (which in general will have discontinuities in its first derivative) is taken as the series of straight lines joining the baseline points and a baseline corrected chromatogram is obtained by baseline subtraction. Peak onset and termination are deduced by considering the first points to the left and right of the local maximum which fall below a threshold corresponding to a user-defined multiple of the noise level. This allows a quantitation baseline to be drawn, which is then subtracted for all points deemed to be part of the peak before they are summed to give a peak area.

For fitting methods, many functions have been put forward to model chromatographic peaks. One of the most successful (although it only takes into account peak tailing) is the exponentially modified Gaussian (EMG) function.[4] As used here, this has the form

$$G(s) = \begin{cases} \dfrac{a_0}{2a_3}\exp\left(\dfrac{a_2^2}{2a_3^2}+\dfrac{a_1-s}{a_3}\right) \\ \quad\times\left[1-erf\left(\dfrac{-s+a_1}{\sqrt{2}\,a_2}+\dfrac{a_2}{\sqrt{2}\,a_3}\right)\right]+B \quad s<a_1+\dfrac{a_2^2}{a_3} \\ \dfrac{a_0}{2a_3}\exp\left(\dfrac{a_2^2}{2a_3^2}+\dfrac{a_1-s}{a_3}\right) \\ \quad\times\left[1+erf\left(\dfrac{+s-a_1}{\sqrt{2}\,a_2}-\dfrac{a_2}{\sqrt{2}\,a_3}\right)\right]+B \quad s\geqslant a_1+\dfrac{a_2^2}{a_3} \end{cases} \tag{1}$$

where a_0 is the intensity, a_1 the point of elution, a_2 the width parameter of the Gaussian, a_3 a parameter defining the distortion and B a background value. The error function ($erf(x)$) is defined as

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)\,du \tag{2}$$

It is important to note that in each of the preceding examples the areas of the minor and major peaks are calculated independently.

In contrast to these approaches, the method proposed here for the determination of isotopomer ratio depends only upon the assumption that the chromatographic peak shape is the same for all the isotopomers of any species to be measured. No attempt is made to define the peak shape function $P(s)$ (s is the mass spectrometer scan number, related to time). This function is superimposed upon a background function $B_M(s)$, so that the expression describing the recorded mass spectrometer output for the major isotopomer $M(s)$ may be written as

$$M(s) = P(s) + B_M(s) \tag{3}$$

To derive a similar expression for the other isotopomers, it must be remembered that because masses are not measured simultaneously, but sequentially, there may be a small apparent delay or lag which should be incorporated into the expression. There may also be a slight difference in retention time for chromatographic reasons. The expression for a peak with isotopomer ratio ρ is therefore

$$J(s) = \rho P(s+\delta) + B_J(s+\delta) \tag{4}$$

where δ is the lag and B_J the function describing the background. Expanding this as a Taylor series about s gives

$$J(s) = \rho\left[P(s) + \delta P'(s) + \frac{\delta^2}{2}P''(s)\right] + \left[B_J(s) + \delta B_J'(s) + \frac{\delta^2}{2}B_J''(s)\right] \tag{5}$$

where the series has been truncated after the third term. This can now be combined with the equation for $M(s)$ to obtain

$$J(s) = \rho\left[M(s) + \delta M'(s) + \frac{\delta^2}{2}M''(s)\right] + [B_J(s) - \rho B_M(s)] + \delta[B_J'(s) - \rho B_M'(s)] + \frac{\delta^2}{2}[B_J''(s) - \rho B_M''(s)] \tag{6}$$

Making use of the approximations

$$M'(s) = [M(s+1) - M(s-1)]/2 \tag{7}$$

and

$$M''(s) = M(s+1) - 2M(s) + M(s-1) \tag{8}$$

and assuming that backgrounds approximate to a straight line, $B(s) = Ps + Q$, we obtain

$$J(s) = \frac{\rho\delta}{2}(\delta+1)M(s+1) + \rho(1-\delta^2)M(s) + \frac{\rho\delta}{2}(\delta-1)M(s-1) + (P_J - \rho P_M)s + (P_J + \delta Q_J - \rho P_M - \rho\delta Q_M) \tag{9}$$

In practice, the digitized values of $J(s)$ are least squared fitted to $M(s+1)$, $M(s)$, $M(s-1)$ and s, the sum of the

coefficients of the fit to the first three terms giving the isotopomer ratio ρ, whilst the lag may be calculated from the sum of the first and third coefficients divided by ρ.

An equivalent approach is to use matrix algebra. If a total of n scans are to be used then the equation corresponding to Eqn (9) is

$$\mathbf{J} = \boldsymbol{M} \cdot \mathbf{P} \tag{10}$$

where $\mathbf{J}$ is the $(n-2)$ row vector of the data for the minor peak. This vector omits the first and the last data recorded. $\boldsymbol{M}$ is a 5 column by $(n-2)$ row matrix, the first column of which is the digitized data for the major peak starting from the first reading and extending to the $(n-2)$th reading. The second column is the same data but starting with the second reading and the third column is the same, but this time beginning with the third reading and extending to the nth. The fourth column of $\boldsymbol{M}$ is the scan number and every element of the fifth column is equal to one. $\mathbf{P}$ is a vector of the best estimates of the five parameters of the fit. The solution of this equation for $\mathbf{P}$ is given by

$$\mathbf{P} = (\boldsymbol{M}^{+}\boldsymbol{M})^{-1}\boldsymbol{M}^{+}\mathbf{J} \tag{11}$$

Calculations

All the calculations were performed using the Excel spreadsheet package (Microsoft). This package is supplied with functions for multi-dimensional least-squares fitting and also with functions for matrix manipulation, allowing either of the two equivalent approaches to be used for fitting using the proposed method.

For the calculation of the EMG profile, Walsh and Diamond[5] used an iterative method, but for better accuracy in the present work an EMG function was written after installing the optional Analysis ToolPak supplied with the Excel software. This function returns a normalized EMG value and is supplied with four arguments corresponding to s, a_1, a_2 and a_3 in Eqn (1). Fitting was accomplished by invoking the built-in Solver( ) function to minimize the sum of the squared residuals between observed and calculated values for the spectra for all values of s by varying the values of a_0, a_1, a_2, a_3 and a static background, B. When performing the fit discretion must be exercised since in some instances there are local minima in the function being optimized which are removed from the global minimum required. Generally the convergence to such a local minimum could be identified by plotting the fitted curve on the same axes as the raw data. If the fit looked poor, then the Solver( ) function was invoked again using different starting points. Further indication of the quality of fit could be had from the rms deviation of the fitted points from the data, but some expertise had to be gained before this parameter could be relied upon.

Gas chromatography/mass spectrometry

GC/MS data were recorded using a Finnigan MAT (Hemel Hempstead, Herts, UK) Incos XL system operating in either the electron ionization (EI) or chemical ionization (CI) mode using methane as a reagent gas. Although this instrument was supplied with a terminal emulator (EM4105 Plus, Diversified Computer Systems, Colorado, USA) which runs under MS-DOS, there was no simple way to get raw data into a form suitable for processing by the Excel package. Accordingly, a short procedure was written under the Finnigan operating system to concatenate the data from the individual scans over the range of interest. These data were then copied from the mass spectrometer data system to the floppy drive of the computer running the emulator. The files so produced had unnecessary text removed and were then processed by a macro written in Excel especially for the purpose. Full details may be obtained from the authors.

RESULTS

In order to illustrate the functioning of the algorithm, a set of pure data was generated. For the purposes of this illustration, the chromatographic peak was represented by a Lorentzian:

$$P(s) = a_1 \frac{a_2^2}{(a_2^2 + s^2)} \tag{12}$$

where a_1 is a normalization factor, equal to $(a_2\,\pi)^{-1}$ for unit area under the peak, and a_2 the width parameter. Table 1 shows the effectiveness of the method in comparing two peaks which are identical except that one is displaced by 0.5 units along the s-axis. This corresponds to the rather artificial case of measuring two channels with an actual isotopomer ratio of unity. For the construction of this table a_2 was set equal to 4 and data were generated for integral values of s in the range -20 to $+20$.

The first three lines of Table 1 show the effect of increasing the number of terms of the Taylor expansion used in the method. For this calculation the first and second derivatives of the expression were evaluated analytically and values of ρ and δ deduced for expansions terminating at the zeroth, first and second derivative terms. The final line of the table shows the results obtained from the full fit as described above, which makes use of approximations for the calculation of the derivatives. This is the approach which has to be adopted in practice when the function describing the chromatographic peak is unknown.

Table 1 demonstrates three important points. First, there is a considerable improvement in the fit as the first and then the second derivative is added to the fit.

Table 1. Investigation of the method as a function of the truncation of the Taylor series used in its derivation[a]

Order of truncation	Ratio, ρ	Lag, δ	r	SS
Zeroth	0.99266	N/A	0.98524	3.08×10^{-4}
First	0.99266	0.500	0.99988	2.47×10^{-6}
Second	0.99994	0.496	1.00000	2.79×10^{-8}
Full method	1.00019	0.512	0.99999	2.23×10^{-7}

[a] Data for Lorentzian curves of equal height, width 4 units and spaced by 0.5 unts ($\rho = 1$, $\delta = 0.5$). The final row, Full method, uses the proposed approximations for the first and second derivatives rather than the exact values as computed for the other three rows.

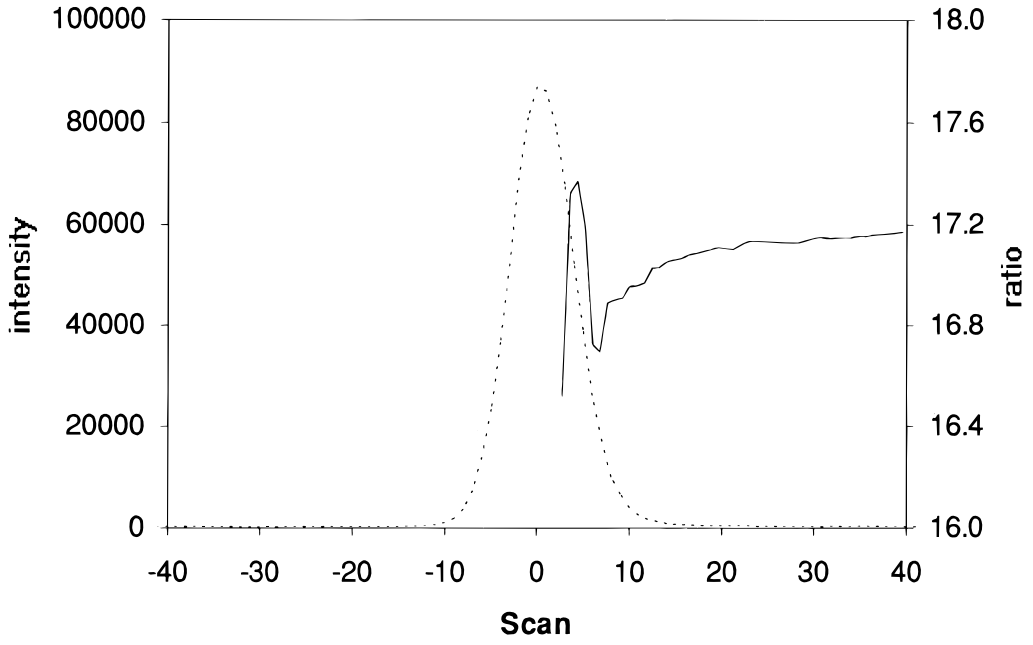

Figure 1. Effect of peak truncation on isotopomer calculation for methyl myristate. The left-hand axis and dotted line show the recorded intensity for the major ion. The right-hand axis and solid line show the isotopomer ratio obtained by considering only that range of scans each side of the peak maximum.

However, it is important to note that the estimate of the ratio of the intensities of the two envelopes is not improved until the second derivative is added. Finally, the approximations made for the calculation of the derivatives from the experimental curves cause less disruption of the fit than neglecting the derivative terms altogether.

It is well known that symmetric functions such as the Gaussian and Lorentzian curves are not ideal for observed chromatograms; in order to investigate the behaviour of the proposed method under more realistic conditions, GC/MS of methyl myristate was performed using a Finnigan MAT Incos XL instrument and selected ion monitoring. To eliminate the enhancement of the M + 1 fragment due to ion–molecular effects common in the EI mass spectra of fatty acid methyl esters,[6] CI was employed.

Whether quantitation is performed by summation or by some form of lineshape analysis, it is important to consider the contributions made by the wings of the peak. However, there must be some point in the peak wings beyond which any further data are ignored. This is called peak truncation and can have dramatic effects on the estimates of peak intensity. Often it is recommended that data should be included for at least four times the peak width and on each side of the peak In instances when chromatograms are crowded (as often happens in physiological studies), this is not possible and significant peak truncation occurs. Figure 1 shows the effects of peak truncation on the calculated isotopomer ratio for the methyl myristate. The isotopomer peaks have been truncated at $\pm$n scans from the maximum of the molecular ion and the isotopomer ratio calculated. This ratio is then plotted against n. Figure 1 indicates that, as with any method, obtaining accurate results critically depends on a good estimate of the baseline. However, it shows that the estimate of isotopomer ratio obtained deviates from its asymptotic value by less than 1% of the estimate when less than two full widths at half-height (FWHH) are considered to each side of the peak.

To allow a comparison to be made between various methods of estimating isotopomer ratios, a set of ten spectra were generated by artificially adding noise to

Table 2. Comparison of peak parameters deduced by various methods[a]

Parameter	Simple summation	Summation	EMG	This method
Area (M)	770536 (1455)	753360 (2952)	759615 (163)	—
Area (M + 1)	136126 (838)	133325 (3226)	128141 (1294)	—
Area (M + 2)	13756 (1268)	16857 (1313)	14884 (1496)	—
(M + 1)/M (%)	17.667 (117)	17.684 (429)	16.869 (170)	17.193 (44)
(M + 2)/M (%)	1.785 (165)	2.237 (170)	1.959 (197)	1.781 (33)

[a] Areas are given in arbitrary units. Numbers in parentheses refer to standatd deviations calculated from simulated data. The simple summation method draws a baseline derived from points remote from the peak. The summation method uses an algorithm as supplied by the instrument manufacturer. The EMG method fits an exponentially modified Gaussian superimposed on a baseline to the data. The column titled 'This method' uses the algorithm porposed in this paper.

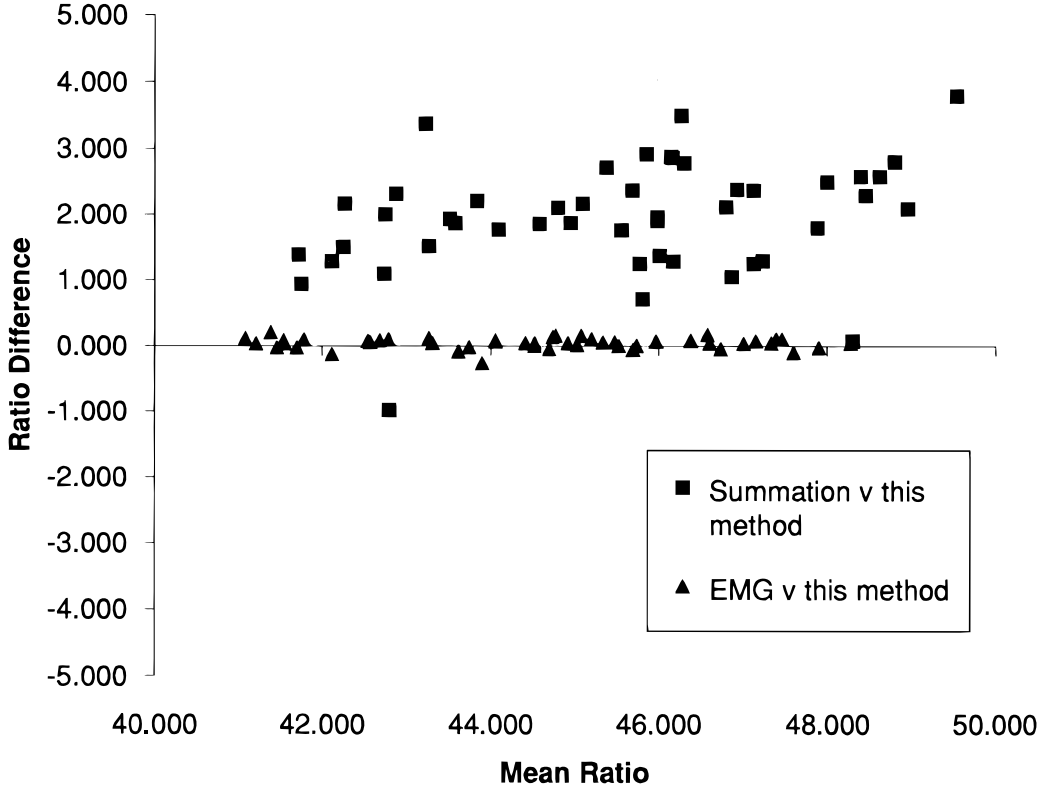

Figure 2. Difference versus average plot of the data for plasma ascorbic acid first isotopomer ratio. The small amount of scatter in the y-direction and the coincidence of the points with the x-axis show that there is very little bias and strong correlation between the method proposed here and the EMG fit.

the recorded spectrum. Instrument noise was estimated from a region of baseline well removed from the peak and sets of random numbers normally distributed about zero and with the same standard deviation as that observed were generated and then added to the experimental spectrum. The isotopomer ratios were estimated by a simple summation method, where the baseline was taken as the straight line joining the mean values of the two extreme sets of ten scans, by a summation method using the algorithms described in the Finnigan MAT

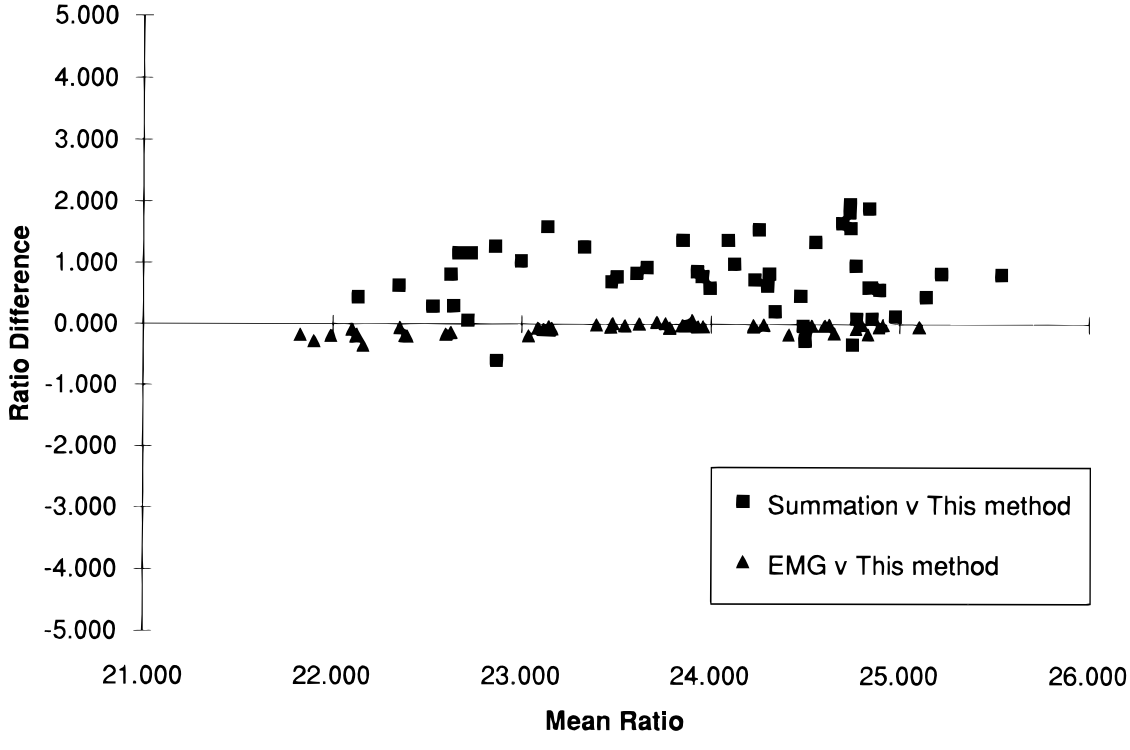

Figure 3. Difference versus average plot of the data for plasma ascorbic acid second isotopomer ratio. The same comments apply concerning the scatter of the data in the y-direction as for the first isotopomer ratio.

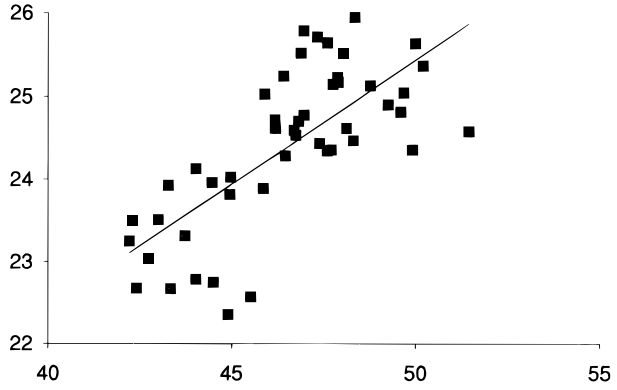

Figure 4. Plot of second versus first isotopomer ratio for plasma ascorbic acid as determined by using the supplied summation algorithm.

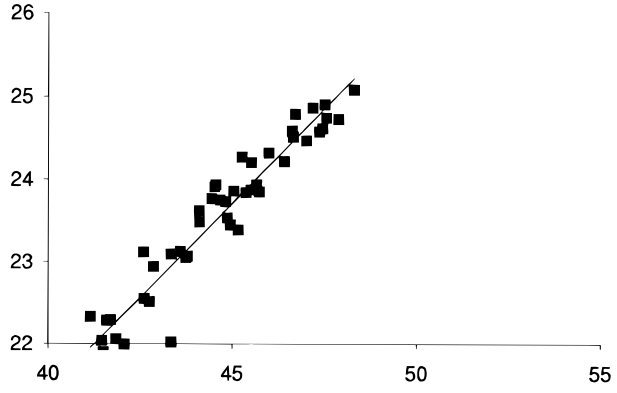

Figure 5. Plot of second versus first isotopomer ratio for plasma ascorbic acid as determined by fitting raw data to the EMG function. This should be compared with the same data for the algorithm supplied by the manufacturer (Fig. 4), which shows much greater scatter.

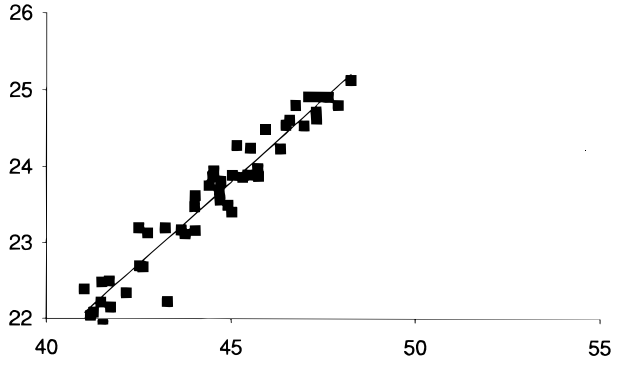

Figure 6. Plot of second versus first isotopomer ratio for plasma ascorbic acid as determined by the proposed method. This should be compared with Figs 4 and 5. The scatter is of the same order as that found for the EMG algorithm and much less than that for the supplied method.

manual,[3] by fitting EMG functions to the individual curves and by the method proposed here. The results are given in Table 2. Note that the last method directly estimates peak ratios and so no assessments of individual peak areas are obtained. It can be seen that the precisions for peak ratios estimated by this method are much superior to those obtained by the other three methods.

It is also of interest to compare the actual values of the ratios obtained from the various methods. For the (M + 1)/M ratio the fitting methods give substantially smaller values than the summation methods. It is difficult to conceive a real GC/MS experiment in which the ratios are precisely known and the generation of dummy data nearly always involves some assumption about lineshape which biases the results to a particular method and so no attempt has been made to judge methods on the 'correctness' of their answers. However, it is worth pointing out that the result obtained from the method proposed in this paper gives values most akin to those obtained from consideration of maximum peak intensities. This is almost bound to be the case given the assumptions made in deriving the algorithm.

As a further test of the method's efficacy, the GC/MS of ascorbic acid in plasma was investigated as part of a metabolic study in man. Full experimental details of this study have been published previously;[7] here we concentrate on the derivation of isotopomer ratios from the GC/MS data and so give only the experimental methods used to follow the time course of a [1-^{13}C] ascorbic acid tracer study.

After formation of the trimethylsilyl derivatives of the sugars in the plasma, the samples were subjected to GC/MS using single ion monitoring of the molecular region under EI. Since both the (M + 1)/M and the (M + 2)/M ratios were to be used in the tracer study, three masses were measured. Each of the 16 samples forming the time course were analysed in triplicate. The ratios were calculated by the summation method, by EMG fit and by the method proposed here.

Figures 2 and 3 show a 'difference versus average' type of analysis[8] for the proposed method versus summation and against EMG fitting for the first and second isotopomer ratios, respectively. It is clear from these figures that there is a much tighter correspondence between the two fitting methods than between the fits and the summation method. It is also clear that there is some systematic discrepancy between the method proposed here and the summation method.

Further evidence for the success of the proposed method is obtained by comparison of the determined (M + 1)/M and (M + 2)/M ratios. It has been shown that for a mixture of two species of differing isotopic composition the (M + 1)/M ratio is linearly related to

Table 3. Statistics for the least-squares fit of second ratio to first ratio for a mixture of ascorbic acid isotopomers as determined by the various methods

Parameter	Summation	EMG	This method
Fractional standard error in slope	14.1	4.3	4.1
Fractional standard error in intercept	18.7	27.6	18.3
Coefficient of determination	0.52	0.92	0.93

the (M + 2)/M ratio regardless of the proportions of the mixture.[7] The corresponding plots are shown in Figs 4, 5 and 6 for the three methods under consideration. Again it can be seen that both the EMG fit and the proposed method greatly reduce the scatter, as is confirmed by parameters from the regression shown in Table 3.

CONCLUSIONS

An expression for an algorithm has been derived which can be used for the determination of isotopomer ratios from recorded GC/MS traces. The algorithm is simple to use and because it does not require any operator input for the determination of backgrounds, etc., it removes the subjectivity encountered in quantitation. No assumptions are made concerning peak shape and so the algorithm is useful in all real cases of peak distortion. It has been tested both on artificial and on real data and has been shown to be superior to the summation type of algorithms typically supplied with the instruments. Comparisons have also been made with fitting to an exponentially modified Gaussian function where the proposed algorithm has proved at least as good in all cases, yet since the fitting is essentially linear the computations involved are much simpler and do not require iterative optimization of the fitting parameters. The ease with which the isotopomer ratios can be measured with increased precision extends the usefulness of the tracer methodology in metabolic studies, particularly when dealing with large body pool sizes where hitherto the doses required might have been considered prohibitive.

REFERENCES

1. J. A. Vogt, T. E. Chapman, D. A. Wagner, V. R. Young and J. F. Burke, et al., *Biol. Mass Spectrum*. **22**, 600 (1993).
2. R. Czoch and A. Francik, *Instrumental Effects in Homodyne Electron Paramagnetic Resonance Spectrometers*. Ellis Horwood, Chichester (1989).
3. *INCOS 50 MSDS Applications Software Operators Manual*. Finnigan MAT, Hemel Hempstead.
4. N. S. Wu, C. P. Cai and Y. Yang, *Chromatographia* **30**, 220 (1990).
5. S. Walsh and D. Diamond, *Talanta* **42**, 561 (1995).
6. R. M. Silverstein, G. C. Bassler and T. C. Morrill, *Spectrometric Identification of Organic Compounds*. Wiley, New York (1974).
7. L. J. C. Bluck, A.P. Izzard and C. J. Bates, *J. Mass Specrrom*. **31**, 741 (1996).
8. J. M. Bland and D. G. Altman, *Lancet* **345**, 1085 (1995)

