ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Chromatography A, 1019 (2003) 31–42

JOURNAL OF CHROMATOGRAPHY A

www.elsevier.com/locate/chroma

Chemometric approach for the resolution and quantification of unresolved peaks in gas chromatography–selected-ion mass spectrometry data

Carlos G. Fraga*

Department of Chemistry, 2355 Fairchild Dr., Suite 2N225, United States Air Force Academy, Colorado 80840-6230, USA

Abstract

A semiautomated and integrated chemometric approach is presented for the resolution and quantification of unresolved target-analyte signals in gas chromatography-selected-ion monitoring (GC-SIM) data collected using scanning mass spectrometers. The chemometric approach utilizes an unskewing algorithm and two multivariate chemometric methods known as rank alignment and the generalized rank annihilation method (GRAM). The unskewing algorithm corrects the retention-time differences within a single GC-SIM data matrix caused by using a scanning mass spectrometer. Rank alignment objectively corrects the run-to-run retention-time difference between a sample GC-SIM data matrix and a standard addition GC-SIM data matrix. GRAM analysis uses the sample and standard addition data matrices to mathematically resolve and quantify the target-analyte signal(s). The resolution and quantification of severely unresolved target-analyte signals are demonstrated using GC-SIM data obtained from conventional heart-cut two-dimensional gas chromatography with mass spectrometric detection. In addition, the GC-SIM data is used to demonstrate the result of chemometric analysis when the absence of a target-analyte signal is obscured by interference. Chemometric analysis is shown to unambiguously detect an analyte based on its resolved mass chromatograms in situations where the traditional approach of measuring peak height fails to positively detect it. The predicted analyte concentrations are within 8% of the reference concentrations.
© 2003 Published by Elsevier B.V.

Keywords: Chemometrics; Generalized rank annihilation method; Curve resolution; Unresolved peaks

1. Introduction

The unambiguous detection and quantification of specific compounds is the ultimate goal for many chemical analyses. In complex mixtures, analyte detection and quantification can be very challenging because of the likelihood of signal interference from mixture constituents. Gas chromatography with detection by mass spectrometry (GC-MS) is frequently used for the analysis of volatile and semi-volatile compounds found in complex mixtures. The mass spectrometer is typically a low-resolution quadrupole or mass selective detector (MSD). An MSD is well suited for the analysis of trace target analytes in complex mixtures because when used in the selected-ion monitoring (SIM) mode, it can be highly sensitive, selective, and quantitative. In SIM mode, the MSD

* Fax: +1-719-333-2947.
E-mail address: carlos.fraga@usafa.af.mil (C.G. Fraga).

0021-9673/$ – see front matter © 2003 Published by Elsevier B.V.
doi:10.1016/S0021-9673(03)01329-3

collects only the signals for ions having masses of interest rather than masses that span a wide range as in full scan mode. The MSD spends more time accumulating signal at a specific mass with SIM, thereby giving SIM better sensitivity than full scan mode for analytes producing ions having specific masses. SIM also provides GC with selective analyte detection by revealing the signal of a target analyte(s) in the presence of co-eluting compounds that do not share the same ionic masses. In practice, SIM can entail the monitoring of a single ion or multiple ions from a target analyte [1]. In this paper, three ions from each analyte are monitored in SIM mode in order to have good sensitivity and enough characteristic ions for confident analyte detection. The detection of an analyte is substantiated if the relative abundance of the selected ionic masses, i.e. the analyte's SIM mass spectrum, matches that of a pure analyte standard. A SIM mass spectrum is actually a vector of signal intensities that is generated by each scan of the MSD during a GC-SIM run. Hence, the GC-SIM data is a matrix consisting of several SIM mass spectrum vectors. After an analyte has been positively detected by its SIM mass spectrum, a calibration plot can be used to quantify the detected analyte based on peak height or area from either a single mass channel or total ion-current plot.

In complex mixtures, the detection and quantification of a target analyte is not always possible with GC-SIM data because signal interference can conceal the analyte signal. Without a resolved SIM mass spectrum, an analyte cannot be positively detected. Conversely, signal interference can also conceal the absence of an analyte signal. Several authors have reported methods for extracting pure mass spectra from GC-MS data [2–10]. Recently, a method for mass spectrum extraction of GC-MS data has been incorporated into an automated software program developed by the National Institute of Standards called Automated Mass Spectral Deconvolution and Identification System (AMDIS) [11]. AMDIS has been used for the resolution of unresolved mass spectra from the GC-MS analysis of complex mixtures [12,13]. The LECO Corporation has also developed a software program used to resolve overlapped signals from GC-time-of-flight mass spectrometry (GC-TOFMS) data [14–16]. Both AMDIS and the LECO method work well, but have one main limitation: they require at least one unique mass fragment for each overlapped component in order to resolve their mass spectra. This requirement is not met with the GC-SIM data in this paper and is rarely met with unresolved GC-SIM data from complex mixtures. However, factor analysis methods such as evolving factor analysis (EFA) do not have this specific requirement. EFA has been used to resolve overlapped signals from GC-MS data [5]. Unfortunately, factor analysis methods that use a single data matrix like EFA are inherently not as robust as those that use two or more matrices [17,18]. The chemometric approach described in this paper does not have the limitations listed above because it uses a bi-matrix factor analysis method known as the generalized rank annihilation method (GRAM). However, a potential problem with GRAM analysis is that it, unlike single-matrix methods, requires GC-SIM data from a sample having a known amount of the target analyte. GC-SIM data obtained from a single calibration standard or a standard-addition sample can satisfy this requirement.

In order to use GRAM, the GC-SIM data needs to be preprocessed in order to remove retention-time shifts. The chemometric approach discussed in this paper utilizes two methods to correct retention-time shifts. An unskewing algorithm developed by Pool et al. [19] is used to correct retention-time shifts caused by scanning mass spectrometers (e.g. MSD), while a chemometric method known as rank alignment is used to correct run-to-run retention time shifts. Rank alignment and GRAM have been used previously on parallel-column GC-TOFMS data [20]. The work in this current paper extends the utility of rank alignment and GRAM to GC-SIM through the use of the unskewing algorithm. In addition, this current paper demonstrates how signal resolution of GC-SIM data by GRAM analysis can be used to unambiguously detect a target analyte in the presence of substantial interference. While other chemometric methods have been applied to GC-SIM data [21,22], to the author's knowledge, this paper is the first to specifically address signal resolution of GC-SIM data by chemometric means. The chemometric approach employed here is fast, easy, and semiautomated, and works well for extracting and quantifying poorly resolved target-analyte signals in GC-SIM data.

1.1. GC-SIM data

Rank alignment and GRAM are used for signal alignment, resolution, and quantification of overlapped GC-SIM signals. Rank alignment is used to objectively align unresolved signals in common between two GC-SIM data matrices of equal size. The two GC-SIM data matrices, a sample data matrix and a standard data matrix, are aligned along the GC column axis by rank alignment before signal resolution and quantification by GRAM. The sample data matrix is that part of GC-SIM data containing a signal suspected to be from a target analyte and originating from an unknown mixture. The standard data matrix contains the signal of the target analyte originating from the unknown mixture spiked with a known amount of target analyte, i.e. standard addition.

1.2. Bilinear data and unskewing algorithm

GC-SIM data matrices must be bilinear or approximately bilinear prior to rank alignment and GRAM analysis. A bilinear data matrix is a data matrix in which each component signal in the data matrix can be individually represented as the product of two vectors. For a bilinear GC-SIM data matrix, each component signal equals the product of a single vector representing the component's GC profile and a single vector representing the component's SIM mass spectrum. The entire bilinear GC-SIM data matrix equals the product of a matrix composed of every component's GC-profile vector and a matrix composed of every component's SIM mass spectrum vector. Fig. 1 illustrates the concept of bilinear data as it pertains to a simulated bilinear GC-SIM data matrix having two overlapped analyte signals.

Bilinear GC-SIM data, like that in Fig. 1, do not originally exist because GC-SIM data are skewed. Skewing occurs because the GC separation produces a changing component concentration in the ion source of the MSD during the measurement of one scan. Fig. 2A depicts GC-SIM data in which skewing has resulted in a single component having mass chromatograms whose peaks have different retention times. This data is clearly not bilinear because one GC-profile (peak shape, width, and retention time) cannot accurately describe all three mass-chromatogram peak profiles. Fortunately, skewed GC-SIM data can be corrected by applying an unskewing algorithm to make the GC-SIM data approximately bilinear. Fig. 2B depicts the GC-SIM data after unskewing using the algorithm described by Pool et al. [19]. The unskewing algorithm corrects the signal intensity for each mass using an equation requiring the measured intensity and the time needed to go from the start of the scan to each mass. The equation is based on the assumption that over a period of three successive scans a parabola describes a GC peak shape. This algorithm has been

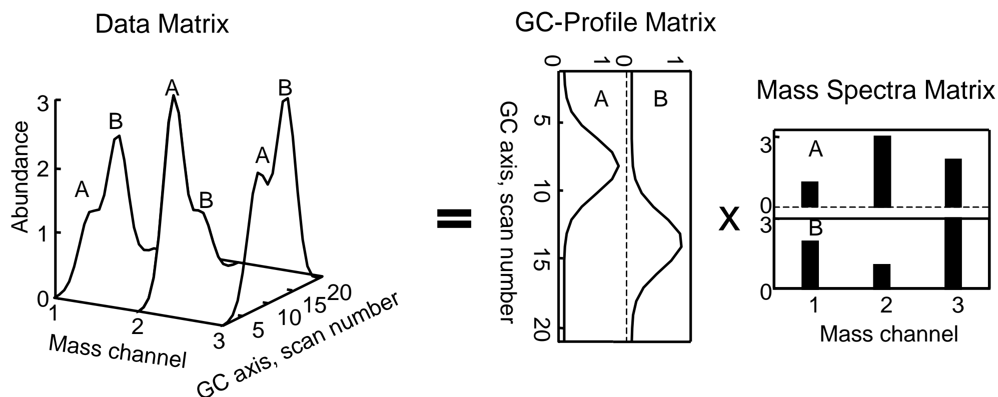

Fig. 1. A simulated bilinear GC-SIM data matrix, in the form of a 3-D plot, equals the product of the GC-profile matrix and the mass spectra matrix. The data matrix is composed of three vectors with each vector representing a mass chromatogram, i.e. the GC signal for a specific mass. The GC-profile matrix contains each component's (A and B) GC-profile vector while the mass spectra matrix contains each component's mass spectrum vector.

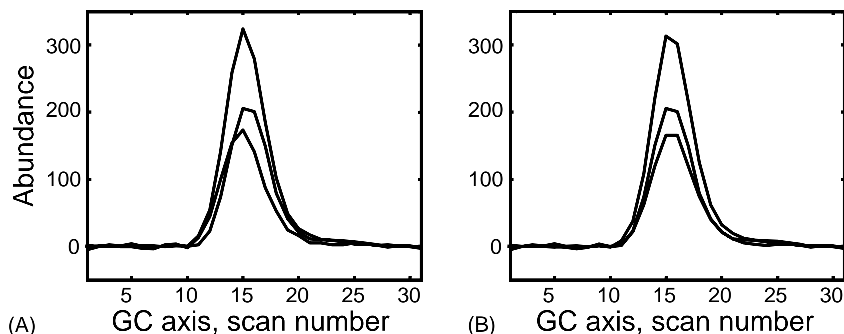

Fig. 2. (A) Raw GC-SIM data of a single analyte depicted as an overlay of three mass chromatograms. (B) The GC-SIM data depicted in (A) after it was unskewed.

successfully used by AMDIS on unresolved triplet peaks each with a base width of seven scans and with less than a scan of separation [11]. As seen in Fig. 2B, the unskewed GC-SIM data is sufficiently bilinear because one GC-profile vector can closely describe all three peak profiles and one SIM mass spectrum can describe each peak's relative height. The problems associated with not unskewing GC-SIM data prior to rank alignment and GRAM analysis are demonstrated later.

1.3. Rank alignment

Once GC-SIM data are unskewed, rank alignment is used prior to GRAM analysis to ensure the common GC-SIM signals between a sample data matrix and standard data matrix are aligned along the GC column axis. Rank alignment has been used to correct run-to-run retention-time shifts in data from hyphenated techniques such as GC-TOFMS, comprehensive two-dimensional gas chromatography, and liquid chromatography with absorbance detection [20,23–26]. Rank alignment does not correct retention-time shifts that affect the resolution between overlapped components. It assumes the resolution between overlapped peaks is constant between the sample and standard data matrices. Rank alignment works by incrementally shifting all the signals in the standard data matrix along the GC column axis. The percent residual variance at each signal shift is then calculated for the matrix formed by adjoining the standard matrix to the sample matrix along their chromatographic dimension. The shift producing a minimum percent residual variance is the shift along the GC column axis for the standard matrix that aligns it to the sample matrix. The percent residual variance is based on the estimated number of components in the sample matrix. Several methods exist for estimating the number of components [27–30]. The number of components inputted for rank alignment is the same as that inputted for GRAM analysis.

Both rank alignment and GRAM are fairly insensitive to inputting a number of components slightly greater (e.g. 1 or 2) than the actual number. This is a frequently reported advantage of GRAM [31]. However, both methods fail if the number of components inputted is less than the actual number.

1.4. GRAM analysis

GRAM is designed to resolve the bilinear signals of analytes that vary in concentration between a sample data matrix and a standard data matrix. It also provides the standard-to-sample concentration ratio for the resolved analytes. Different versions of the GRAM algorithm are discussed in detail by several authors [31–35]. The GRAM algorithm used in this paper comes from Wilson et al. [34]. Its only required input is an estimate of the number of components present in the sample matrix. The sample and standard matrices must meet five requirements. First, as stated earlier, each of these equal-sized data matrices needs to be bilinear. Second, these bilinear data matrices as a group must also be trilinear. Trilinear means the bilinear signals in common between the sample and standard matrices are identical, excluding signal intensity. Trilinear GC-SIM data will have a GC profile (i.e.

peak shape, width, and retention time) and a SIM mass spectrum that are constant for a compound present in both matrices. As previously shown, unskewing addresses the bilinearity requirement. On the other hand, the trilinearity requirement is addressed by rank alignment and standard addition. Specifically, rank alignment ensures analytes in common have the same retention time while standard addition ensures an analyte's peak shape and width are constant by eliminating chemical matrix effects [23]. The variability of an analyte's mass spectrum is not mathematically corrected but is usually small enough for successful GRAM analysis. The third requirement is that a target analyte cannot have the same SIM mass spectrum or the same GC profile as another component. Isomers having almost identical mass spectra that perfectly or very closely co-elute will not meet this requirement. The end result will likely be a false positive. Fourth, a target analyte(s) cannot perfectly covary with other components in terms of concentration between the sample and standard data matrices. The standard addition approach ensures compliance with this requirement. The final requirement is that the total number of unique component signals in the sample and standard data matrices cannot exceed the smallest dimension of the data matrices. A component signal is considered unique if it has a unique GC profile and a unique mass spectrum. This last requirement stems from the fact that the smallest dimension of a data matrix equals the maximum number of signals GRAM analysis can output. The number of mass channels is typically the smallest dimension of a GC-SIM data matrix. Because an offset baseline is considered a unique component signal, a linear background correction is used on all data matrices to help meet the above requirement.

2. Experimental

2.1. Data collection

GC-SIM data were provided from the Midwest Research Institute (MRI, Kansas City, MO, USA). The GC-SIM data were acquired from a conventional heart-cut two-dimensional (2-D) GC-SIM (GC/GC-SIM) analyzer set to selectively monitor three mass channels. The GC/GC-SIM analyzer consisted of two GC columns of different selectivity connected in series by a valve. A heart-cut containing a target analyte was diverted from the first column eluent by a valve to the second GC column for further separation and subsequent detection by an MSD. No more than two heart-cuts were made for each GC/GC-SIM analysis. Each heart-cut had a width slightly larger than the base width of the target's first column peak and was centered at the target analyte's retention time on the first column. The target analyte's exact retention time was determined by standard addition. Each heart-cut was cryogenically re-focused prior to separation on the second GC column in order to get sample concentration and higher chromatographic efficiency.

Two sets of GC-SIM data were used in this study. Each set of data was a collection of GC-SIM data files acquired during the GC/GC-SIM analysis of one of two test solutions. Each test solution was collected from the liquid solvent extraction of a separate environmental sample. Solution 1 was analyzed for triethyl phosphate (TEP) and solution 2 was analyzed for 1,4-dithiacyclohexane (DCH). The mass channels monitored for TEP had mass-to-charge (m/z) ratios of 81, 99, and 155. The m/z ratios for DCH were 46, 61, and 120. All mass channels selected were chosen because of good sensitivity and selectivity for the target compound. Each of the two GC-SIM data sets (solutions 1 and 2) included two GC-SIM data files: one for the GC/GC-SIM analysis of the test solution and the other from the GC/GC-SIM analysis of the test solution spiked with a known amount of target compound.

Each GC-SIM data set also included several GC-SIM data files from replicate GC/GC-SIM analyses of a standard solution containing a known amount of the target analyte (TEP or DCH). The replicate analyses of each standard solution were completed shortly before and after the analysis of the given test solution. The replicate GC/GC-SIM analyses of each standard solution were used to provide 95% confidence limits for the ion-abundance ratios of each target compound. The confidence limits were based on a two-sided t-distribution. In addition to the GC-SIM data, chromatograms and quantitative results based on a calibration curve were obtained for the analysis of solution 2 by a GC-flame photometric detector (GC-FPD). The GC-FPD results are used only as a reference for those provided by the chemometric analysis of GC-SIM data from solution 2.

2.2. Data analysis

Each GC-SIM data file was rearranged into a matrix of data using MassTransit 2.6.1 (Palisade Corp., Newfield, NY, USA) and Microsoft Excel 97. Data handling and chemometric analysis of the GC-SIM data matrices were all completed using Matlab 5.2 (The Mathworks, Inc., Natick, MA, USA). Chemometric analysis was accomplished using a Matlab program written in-house that sequentially and automatically applied linear-background correction, signal unskewing, rank alignment, and GRAM analysis on the data matrices. The Matlab program typically took 0.1 s to process and output the results using a Pentium 2.0 GHz computer with 512 MB RAM. The only required inputs for the program were two data matrices (sample and standard), the estimated number of components, and the expected maximum retention-time shift in number of scans. The program determined which GRAM component was believed to be the target analyte. Because standard addition of the target analyte was performed, the GRAM component with the largest standard-to-sample concentration ratio was automatically considered the target analyte. The remaining GRAM components were considered interferents and their signals were added to make one signal.

2.3. Simulations

Two simulations, A and B, were produced to demonstrate and evaluate the chemometric analysis of GC-SIM data under known conditions that could be encountered when analyzing target compounds in complex environmental mixtures. Each simulation consisted of the chemometric analysis of a sample data matrix and a standard data matrix. A maximum shift of two scans and an expected number of components of three were used in both simulations.

In simulation A, the sample data matrix is 17-by-3 in size and contains a TEP signal overlapped with two distinct component signals. The TEP signal was originally resolved chromatographically but was artificially overlapped with the signals from the other two components. This was done by adding together three 17-by-3 GC-SIM sub-matrices. Each sub-matrix was comprised of three mass chromatograms each with 17 data points. The first sub-matrix contained the chromatographically resolved GC-SIM signal of TEP from the analysis of a 10 ng/ml TEP solution. The second sub-matrix originated from the GC/GC-SIM analysis of solution 1, and it contained a single GC profile present in all three mass channels. The third sub-matrix contained the resolved GC-SIM signal of DCH from the analysis of a 44 ng/ml DCH solution. The two interferents in the sample data matrix have resolutions (Rs) of 0.1 and 0.3 with TEP. The standard data matrix was formed by adding the sample data matrix and the same TEP sub-matrix used previously. The standard data matrix was then shifted by one scan to simulate a run-to-run retention time shift.

In simulation B, the sample data matrix was created in the same way as in simulation A, although the TEP sub-matrix was not included. The standard data matrix was formed by adding the sample data matrix and the TEP sub-matrix. It was then shifted by one scan.

2.4. Non-simulated application

A real-world application was performed through the chemometric analysis of GC-SIM data collected from the GC/GC-SIM analyses of solution 2. The application involved the chemometric analysis of a sample data matrix and standard data matrix each 10-by-3 in size. A maximum shift of zero and an expected number of components of three were inputted into the Matlab program. The standard data matrix contains the unresolved signal of DCH originating from its standard addition into an aliquot of solution 2. The sample data matrix contains the unresolved GC-SIM signal of DCH originally in solution 2. A known amount of DCH had been spiked into the environmental sample prior to solvent extraction. GC-FPD analysis of the extract (solution 2) detected the presence of DCH using standard addition, and its concentration was determined using a quadratic calibration curve.

3. Results and discussion

3.1. Chemometric analysis of simulations

Fig. 3A depicts the sample and standard data matrices from simulation A after unskewing. The sample data matrix simulates GC-SIM data from an unknown solution while the standard data matrix simulates GC-SIM data from the standard addition of TEP into

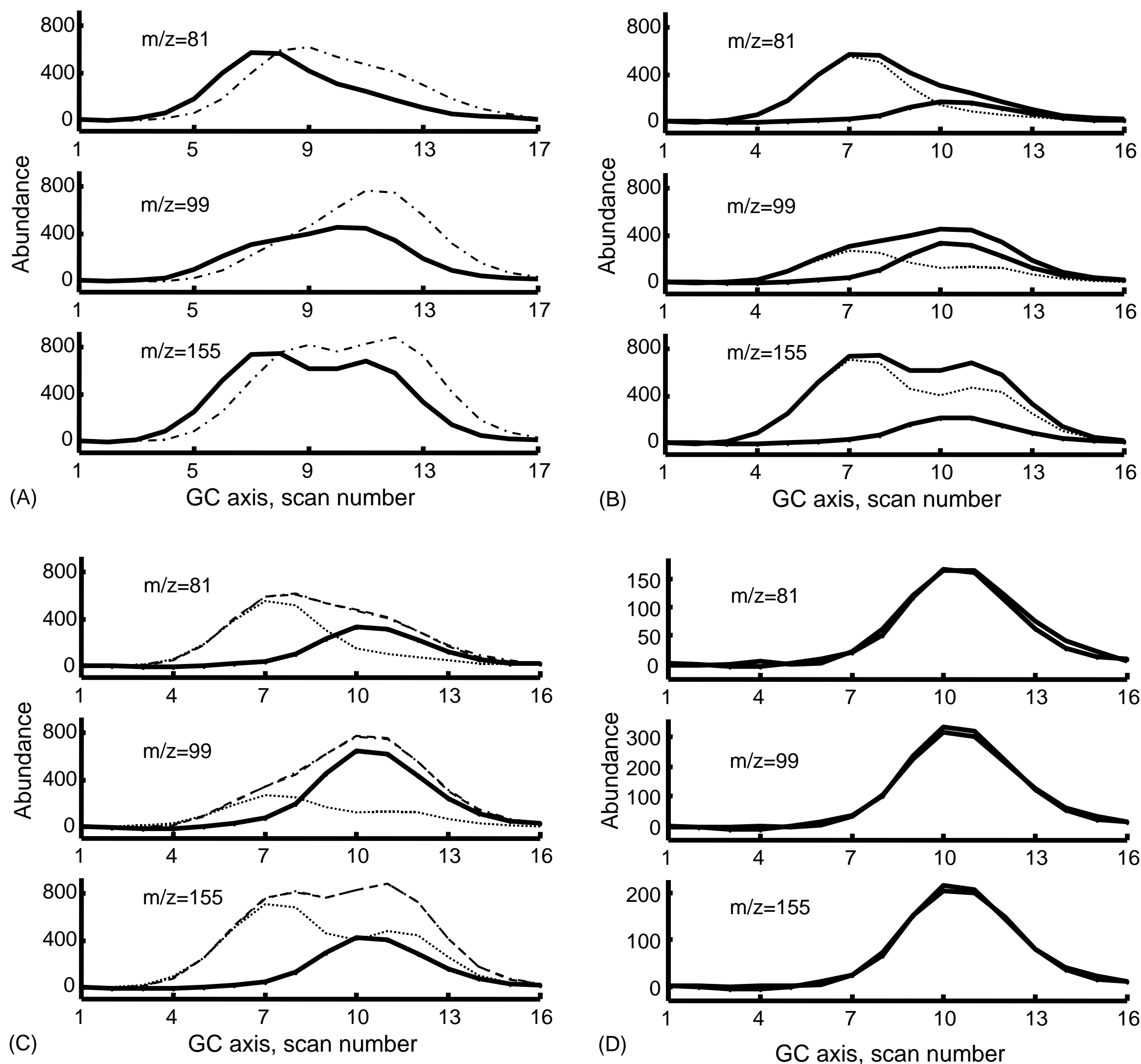

Fig. 3. (A) The sample data matrix (solid line) and standard data matrix (dash dot) from simulation A depicted using three mass-channel plots. Each plot depicts one mass chromatogram from each data matrix. The sample matrix contains the unskewed, unresolved signal of TEP. The standard matrix contains the unskewed, unresolved signal of TEP obtained from the standard addition of TEP into the sample. (B) The sample matrix (solid line) overlaid with the GRAM signals for TEP (solid circles), interference (dots), and reconstructed sample (dash). (C) The standard matrix (dash dot) overlaid with the GRAM signals for TEP (solid circles), interference (dots), and reconstructed standard (dash). (D) The signals for the GRAM resolved TEP (solid circles) and the true unskewed TEP (solid line) for the sample matrix.

the unknown solution. A TEP signal is present in both data matrices but because of signal interference it cannot be detected using the traditional approach. The traditional approach involves measuring the relative ion abundance for each mass channel by either peak area or height. In this paper, peak height is used for all ion abundance measurements because it is less susceptible to inaccurate results from peak overlap. If the relative ion abundance in a data matrix matches that expected for TEP, then TEP is believed to be present. Because of signal interference, valid peak height measurements are impossible for these data matrices. In

this case, the analyst would be forced to conclude that a TEP signal is not detected in either data matrix. Of course, the analyst has the option to perform another run with alternative ions and hope that they are not interfered. Even if they are not interfered, they will normally be of lower abundance and may not be detected because they fall below the detection limit. Another option is for the analyst to use GC-MS deconvolution software provided by the instrument manufacture or another source. However, many GC-MS deconvolution methods have significant drawbacks when it comes to complex GC-SIM data. For example, the deconvolution methods used by LECO and NIST cannot resolve the TEP signal in simulation A because a unique ion is not present. The same fate is likely with the novel method from Gan and Liang [9]. Their method requires the mass spectra for all overlapped components, including the interfering ones, to be in a library. That can be impossible to obtain even with the most comprehensive MS libraries. Curve resolution methods that do not use a standard data matrix, such as any of the single-matrix factor analysis methods, will likely fail when applied to seriously unresolved data [17,18]. This was the case when EFA and multivariate curve resolution from PLS Toolbox [35] were applied to the data from simulation A. Other more complicated single-matrix methods that merge non-iterative and iterative features may work for complex GC-SIM data and should be tested [8]. Fortunately, the GRAM-based chemometric approach used in this paper can easily resolve and quantify the TEP signal.

Rank alignment and GRAM analysis were performed on the sample and standard data matrices seen in Fig. 3A. Fig. 3B depicts the results of GRAM analysis for the sample data matrix. In Fig. 3B, the sample data matrix is displayed with the GRAM signals for TEP and its interference. The sum of both signals produces a GRAM-reconstructed sample data matrix, which is also displayed. Fig. 3B clearly reveals the presence of the three component signals that were purposely overlapped to produce this simulation. They have retention times (scan number) of 7.5, 10.5, and 11.5. Similar GRAM results are depicted for the standard data matrix depicted in Fig. 3C. The ultimate success of GRAM analysis is assessed by comparing the GRAM signal for TEP with the known TEP signal (see Fig. 3D). In Fig. 3D, the bilinear signal representing TEP in the sample data matrix fits remarkably well with the unskewed true signal of TEP. The quantitative accuracy of GRAM analysis is also quite good. The predicted standard-to-sample concentration ratio from GRAM analysis is 1.95. The true value is 2. Rank alignment also accurately

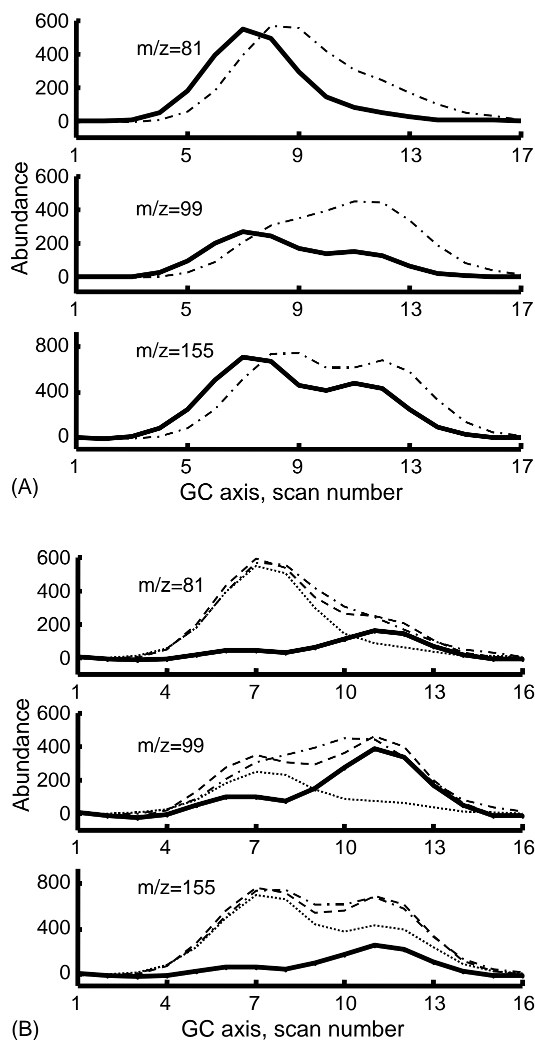

Fig. 4. (A) The sample data matrix (solid line) and standard data matrix (dash dot) from simulation B depicted using three mass-channel plots. The sample matrix is identical to the one from simulation A except the TEP signal is absent. The standard matrix contains the unresolved signal of TEP from the standard addition of TEP into the sample. Each data matrix is unskewed. (B) The standard matrix (dash dot) overlaid with the GRAM signals for the target analyte (solid circles), interference (dots), and reconstructed standard (dash).

determined the true retention-time shift of −1 scan. A negative value shift means the signals in the standard data matrix needed to be shifted to an earlier retention time for proper alignment.

In real-world applications, TEP would be considered present if the ion-abundance ratios for the *m*/*z* combinations of 99/81, 155/81, and 155/99 match those expected for TEP. For the GRAM target analyte from simulation A, the ion-abundance ratios are within the 95% confidence limits of the mean ion-abundance ratios obtained from replicate analyses of a TEP standard solution. Consequently, TEP is considered unambiguously present in simulation A. The detection of TEP would not have occurred in

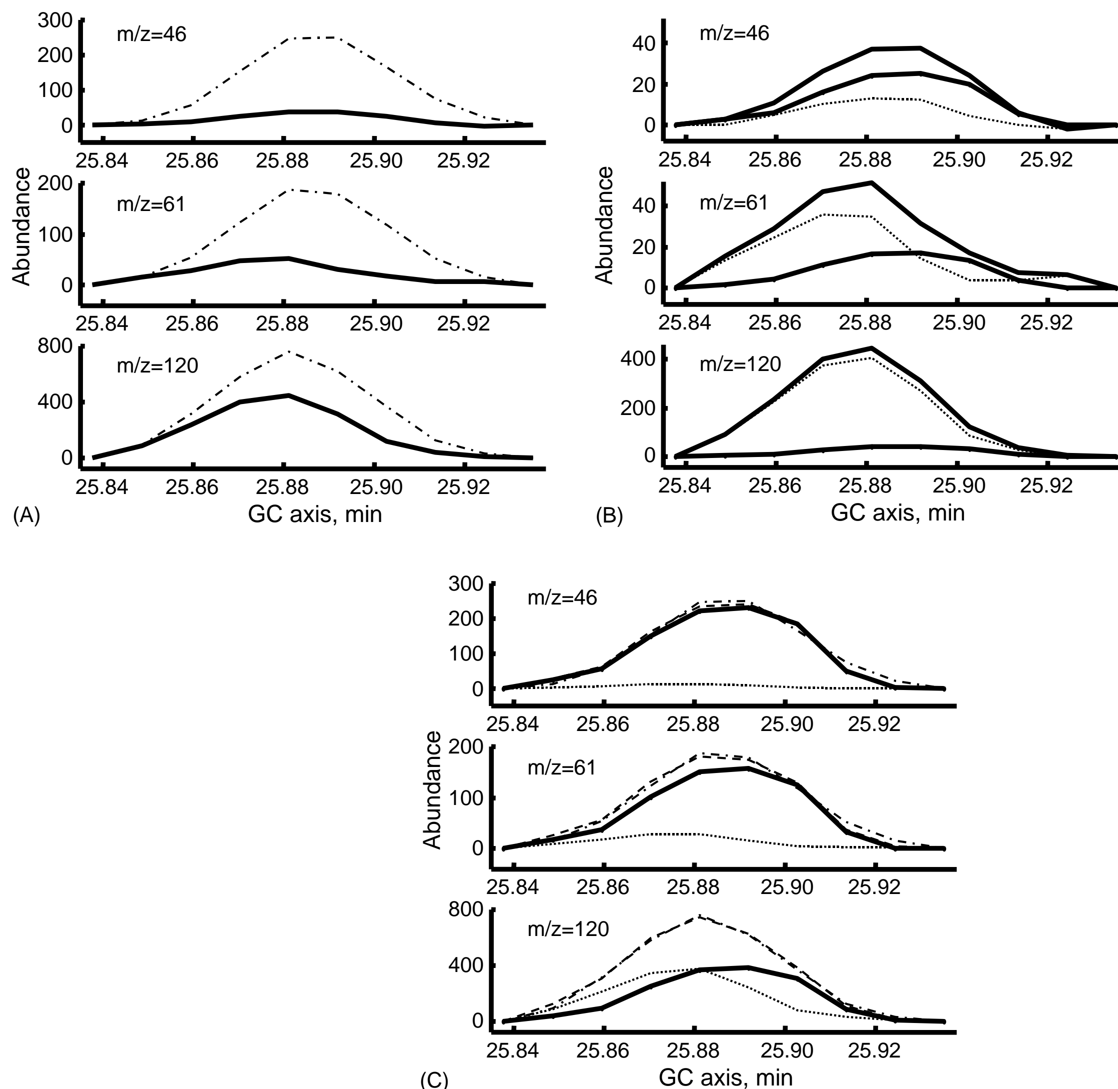

Fig. 5. (A) Mass-channel plots for the unskewed sample and standard data matrices from the non-simulated application. The sample matrix (solid) has an unresolved DCH signal. The standard matrix (dash dot) has the unresolved signal of DCH from the standard addition of DCH. (B) The sample matrix (solid line) overlaid with the GRAM signals for DCH (solid circles), interference (dots), and reconstructed sample (dash). (C) The standard matrix (dash dot) overlaid with the GRAM signals for DCH (solid circles), interference (dots), and reconstructed standard (dash).

this simulation without unskewing the GC-SIM data. Chemometric analysis of simulation A without applying the unskewing algorithm caused two out of three ion-abundance ratios to fall outside the 95% confidence limits for TEP. The negative impact of skewed GC-MS (full scan mode) data on signal resolution by a factor analysis method has been previously noted [5]. For GC-SIM data, skewing is generally more severe because the monitored ions are evenly distributed over an entire mass range.

Simulation B demonstrates chemometric analysis when the target analyte is not present in the sample data matrix. Fig. 4A depicts the sample and standard matrices from simulation B. Because of interference, it is impossible to deduce that TEP is absent in the sample data matrix. TEP is known to be present in the standard data matrix because of standard addition. Fig. 4B depicts the results of GRAM analysis for the standard data matrix. The results for the sample data matrix are not shown for brevity. The bimodal GC profile of the target analyte in Fig. 4B is fundamentally caused by the absence of TEP in the sample data matrix. This unrealistic representation of the target-analyte signal is matched by a poor representation of the standard data matrix. In Fig. 4B the GRAM-reconstructed standard (i.e. the sum of the target-analyte signal and interference signal) does not come close to fitting the standard data matrix. These unrealistic and poor-fitting representations happen because the target analyte is absent in the sample data matrix. Clearly, TEP is not detected by GRAM analysis. This simulation helps demonstrate how a false positive by GRAM analysis is unlikely because performing singular value decomposition on the sample matrix, as part of the GRAM algorithm, ensures that the TEP signal in the standard matrix is never included as a modeled component.

A similar outcome to that depicted in Fig. 4B is also possible when the GRAM requirements discussed in Section 1.4 are not met, even though the target analyte is present. The most likely requirement not to be met is the one requiring at least the same number of mass channels as unique component signals.

3.2. *Chemometric analysis of non-simulated sample and standard*

Fig. 5A depicts the sample and standard data matrices pertaining to solution 2. The sample data matrix has an obvious peak in all three mass channels that closely matches the retention time of the spiked DCH in the standard data matrix. As shown in Table 1, the ion-abundance ratios for the peak in the sample data matrix do not match those expected for DCH even though DCH is present. Moreover, the standard data matrix with the spiked DCH has ion-abundance ratios that do not match those for DCH. Clearly, signal interference inhibits the detection of DCH in both data matrices. Fig. 5B and C show the GRAM results for the sample and standard data matrices, respectively. The GRAM signals from both data matrices are realistic, good-fitting representations of the real data and the ion-abundance ratios for the target analyte fall within the 95% confidence limits for a DCH signal (see Table 1). Hence, GRAM analysis correctly detects the presence of DCH in both data matrices. In addition, the GRAM-calculated DCH concentration of

Table 1
Traditional and chemometric analyses of non-simulated sample and standard

Analysis method	Data matrix	Ion abundance ratios			DCH detected[a]
		46/61	46/120	61/120	
Traditional	Calibration sample	1.49 ± 0.04[b]	0.62 ± 0.04[b]	0.413 ± 0.024[b]	NA
Traditional	Sample	0.682	0.0954	0.139	No
Traditional	Standard	1.31	0.331	0.252	No
Chemometric	Sample & standard	1.47	0.601	0.409	Yes
Chemometric (raw data)	Sample & standard	−1.08	−21.2	19.6	No

NA: not applicable.

[a] DCH is detected if all ion-abundance ratios fall within their respective 95% confidence limits, i.e. 4.3 S.D. from the mean ion-abundance ratio of the calibration sample.

[b] The mean and standard deviation (S.D.) of ion-abundance ratios calculated from three replicate analyses of a 6.2 ng/ml DCH calibration sample.

2.4 ng/ml closely matches 2.6 ng/ml, which is the reference concentration determined by GC-FPD analysis. This successful outcome is not possible without first unskewing the raw GC-SIM data. GRAM analysis on the raw GC-SIM data results in a false negative because unrealistic GRAM results are obtained which include negative ion-abundance ratios that obviously do not match those expected for DCH (see Table 1).

4. Conclusions

The resolution and quantification of unresolved GC-SIM signals can be achieved through the chemometric approach presented. This chemometric approach can in many cases permit the unambiguous detection and quantification of target compounds that are otherwise not detected because of signal interference. In addition to the data presented, it has been used to positively detect and quantify target-compound signals found in unresolved GC-SIM data not presented. However, if this chemometric approach does not detect a target compound, one cannot say for certain the analyte is not present. A non-detect may be due to the data not meeting the discussed requirements. This drawback can be mitigated by utilizing as many significant mass channels as possible such that the critical requirement of having at least the same number of mass channels as unique component signals is met. The number of mass channels will depend heavily on the level of detection required because acquisition time per mass channel using an MSD drops with an increase in mass channels. Hence, full scan analysis can conceivably be used if higher detection limits are acceptable. This integrated chemometric approach should give current GC-SIM methods enhanced capability for detecting and quantify target analytes in very complex mixtures. Moreover, using the signal increase caused by standard addition as a marker for target-analyte location, this chemometric approach can be incorporated into a fully automated program that finds sub-sections of GC-SIM data for subsequent chemometric analysis.

Acknowledgements

The author thanks S. Fitzgibbons from MRI for the analytical data. R.E. Synovec and B.J. Prazen from the University of Washington are also thanked for helpful discussions and advice.

References

[1] R.L. Grob (Ed.), Modern Practice of Gas Chromatography, third ed., Wiley, New York, 1995, p. 337.
[2] W.G. Pool, J.W. de Leeuw, B. van de Graaf, J. Mass Spectrom. 32 (1997) 438.
[3] R.G. Dromey, M.J. Stefik, T.C. Rindfleisch, A.M. Duffield, Anal. Chem. 48 (1976) 1368.
[4] R.F. Lacey, Anal. Chem. 58 (1986) 1404.
[5] L. Roach, M. Guilhaus, Org. Mass Spectrom. 27 (1992) 1071.
[6] M.D. King, G.S. King, Anal. Chem. 57 (1985) 1049.
[7] M.A. Sharaf, B.R. Kowalski, Anal. Chem. 53 (1981) 518.
[8] H. Shen, B. Grung, O.M. Kvalheim, I. Eide, Anal. Chim. Acta. 446 (2001) 313.
[9] F. Gan, Y. Liang, Anal. Sci. 16 (2000) 603.
[10] F. Gong, Y. Liang, Q. Xu, F. Chau, J. Chromatogr. A 905 (2001) 193.
[11] S.E. Stein, J. Am. Soc. Mass Spectrom. 10 (1999) 770.
[12] S.G. Roussis, P.W. Fitzgerald, Energy Fuels 15 (2001) 477.
[13] S. Dagan, J. Chromatogr. A 868 (2000) 229.
[14] M.M. van Deursen, J. Beens, H.G. Janssen, P.A. Leclercq, C.A. Cramers, J. Chromatogr. A 878 (2000) 205.
[15] T. Veriotti, R. Sacks, Anal. Chem. 72 (2000) 3063.
[16] T. Veriotti, R. Sacks, Anal. Chem. 73 (2001) 4395.
[17] R. Tauler, A. Smilde, B.R. Kowalski, J. Chemom. 9 (1995) 31.
[18] A. Juan, S.C. Rutan, R. Tauler, D.L. Massart, Chemom. Intell. Lab. Syst. 40 (1998) 19.
[19] W.G. Pool, J.W. de Leeuw, B. van de Graaf, J. Mass Spectrom. 31 (1996) 213.
[20] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, Anal. Chem. 71 (1999) 1093.
[21] J. Weber, M. Beeg, C. Bartzsch, K.H. Feller, D.C. de la Garcia, M. Reichenbacher, K. Danzer, J. High Resolut. Chromatogr. 22 (1999) 322.
[22] H. van der Voet, W.J. de Boer, W.G. de Ruig, J.A. van Rhijn, J. Chemom. 12 (1998) 279.
[23] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 72 (2000) 4154.
[24] B.J. Prazen, R.E. Synovec, B.R. Kowalski, Anal. Chem. 70 (1998) 218.
[25] B.J. Prazen, C.A. Bruckner, R.E. Synovec, B.R. Kowalski, J. Microcolumn Sep. 11 (1999) 97.
[26] C.G. Fraga, B.J. Prazen, R.E. Synovec, Anal. Chem. 73 (2001) 5833.
[27] N.M. Faber, L.M.C. Buydens, G. Kateman, Anal. Chim. Acta 296 (1994) 1.
[28] N.M. Faber, L.M.C. Buydens, G. Kateman, Chemom. Intell. Lab. Syst. 25 (1994) 203.

[29] K. Faber, B.R. Kowalski, Anal. Chim. Acta 337 (1997) 57.
[30] E.R. Malinowski, Factor Analysis in Chemistry, second ed., Wiley, New York, 1991.
[31] N.M. Faber, L.M.C. Buydens, G.J. Kateman, Chemometrics 8 (1994) 147.
[32] E. Sánchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496.
[33] S. Li, J.C. Hamilton, P.J. Gemperline, Anal. Chem. 64 (1992) 599.
[34] B.E. Wilson, E. Sánchez, B.R. Kowalski, J. Chemom. 3 (1989) 493.
[35] B.M. Wise, N.B. Gallagher. PLS-Toolbox Version 2.0. Eigenvector Research, Inc., Manson, WA, 1998.